DEX-035

BL

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 November 2002 (21.11.2002)

PCT

(10) International Publication Number
WO 02/092785 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/22271

(22) International Filing Date: 13 February 2002 (13.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/268,290	13 February 2001 (13.02.2001)	US
60/268,834	15 February 2001 (15.02.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/268,290 (CIP)
Filed on	13 February 2001 (13.02.2001)
US	60/268,834 (CIP)
Filed on	15 February 2001 (15.02.2001)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SALCEDA, Susana** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **HU, Ping** [US/US]; 108 Palmer Street, San Ramon, CA 94583 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **KARRA, Kalpana** [IN/US]; 6511 Rainbow Drive, San Jose, CA 95129 (US). **CAFFERKEY, Robert** [IE/US]; 849 West Orange Avenue, Apartment 1030, South San Francisco, CA 94081 (US). **SUN, Yongming** [US/US]; Apartment 260, 869 S. Winchester Boulevard, San Jose, CA 95128 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrel P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/092785 A2

(54) Title: COMPOSITION AND METHODS RELATING TO OVARIAN SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of the compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO OVARIAN SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application Serial No. 60/268,290 filed February 13, 2001, and U.S. Provisional Application Serial No. 60/268,834 filed February 15, 2001, which are herein incorporated by reference in their entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

BACKGROUND OF THE INVENTION

Cancer of the ovaries is the fourth-most cause of cancer death in women in the United States, with more than 23,000 new cases and roughly 14,000 deaths predicted for the year 2001. Shridhar, V. et al., Cancer Res. 61(15): 5895-904 (2001); Memarzadeh, S. & Berek, J. S., J. Reprod. Med. 46(7): 621-29 (2001). The incidence of ovarian cancer is of serious concern worldwide, with an estimated 191,000 new cases predicted anually. Runnebaum, I. B. & Stickeler, E., J. Cancer Res. Clin. Oncol. 127(2): 73-79 (2001). Because women with ovarian cancer are typically asypmtomatic until the disease has metastasized, and because effective screening for ovarian cancer is not available, roughly 70% of women present with an advanced stage of the cancer, with a five-year survival rate of ~25-30% at that stage.

Memarzadeh, S. & Berek, J. S., supra; Nunns, D. et al., Obstet. Gynecol. Surv. 55(12): 746-51. Conversely, women diagnosed with early stage ovarian cancer enjoy considerably higher survival rates. Werness, B. A. & Eltabbakh, G. H., Int'l. J. Gynecol. Pathol. 20(1): 48-63 (2001).

Although our understanding of the etiology of ovarian cancer is incomplete, the results of extensive research in this area point to a combination of age, genetics, reproductive, and dietary/environmental factors. Age is a key risk factor in the development of ovarian cancer: while the risk for developing ovarian cancer before the age of 30 is slim, the incidence of ovarian cancer rises linearly between ages 30 to 50, increasing at a slower rate thereafter, with the highest incidence being among septagenarian women. Jeanne M. Schilder et al., Heriditary Ovarian Cancer: Clinical Syndromes and Management, in Ovarian Cancer 182 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001).

With respect to genetic factors, a family history of ovarian cancer is the most significant risk factor in the development of the disease, with that risk depending on the number of affected family members, the degree of their relationship to the woman, and which particular first degree relatives are affected by the disease. Id. Mutations in several genes have been associated with ovarian cancer, including BRCA1 and BRCA2, both of which play a key role in the development of breast cancer, as well as hMSH2 and hMLH1, both of which are associated with heriditary non-polyposis ovary cancer. Katherine Y. Look, Epidemiology, Etiology, and Screening of Ovarian Cancer, in Ovarian Cancer 169, 171-73 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). BRCA1, located on chromosome 17, and BRCA2, located on chromosome 13, are tumor supressor genes implicated in DNA repair; mutations in these genes are linked to roughly 10% of ovarian cancers. Id. at 171-72; Schilder et al., supra at 185-86. hMSH2 and hMLH1 are associated with DNA mismatch repair, and are located on chromsomes 2 and 3, respectively; it has been reported that roughly 3% of heriditary ovarian carcinomas are due to mutations in these genes. Look, supra at 173; Schilder et al., supra at 184, 188-89.

Reproductive factors have also been associated with an increased or reduced risk of ovarian cancer. Late menopause, nulliparity, and early age at

menarche have all been linked with an elevated risk of ovarian cancer. Schilder et al., supra at 182. One theory hypothesizes that these factors increase the number of ovulatory cycles over the course of a woman's life, leading to "incessant ovulation," which is thought to be the primary cause of mutations to the ovarian epithelium. Id.; Laura J. Havrilesky & Andrew Berchuck, Molecular Alterations in Sporadic Ovarian Cancer, in Ovarian Cancer 25 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). The mutations may be explained by the fact that ovulation results in the destruction and repair of that epithelium, necessitating increased cell division, thereby increasing the possibility that an undesried mutation will occur. Id. Support for this theory may be found in the fact pregnancy, lactation, and the use of oral contraceptives, all of which suppress ovulation, confer a protective effect with respect to developing ovarian cancer. Id.

Among dietary/environmental factors, there would appear to be an association between high intake of animal fat or red meat and ovarian cancer, while the antioxidant Vitamin A, which prevents free radical formation and also assists in maintaining normal cellular differentiation, may offer a protective effect. Look, supra at 169. Reports have also associated asbestos and hydrous magnesium trisilicate (talc), the latter of which may be present in diaphragms and sanitary napkins. Id. at 169-70.

Current screening procedures for ovarian cancer, while of some utility, are quite limited in their diagnostic ability, a problem that is particularly acute at early stages of cancer progression when the disease is typically asymptomatic yet is most readily treated. Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 166 (1998); Memarzadeh & Berek, supra; Runnebaum & Stickeler, supra; Werness & Eltabbakh, supra. Commonly used screening tests include bimanual rectovaginal pelvic examination, radioimmunoassay to detect the CA-125 serum tumor marker, and transvaginal ultrasonography. Burdette, supra at 166.

Pelvic examination has failed to yield adequate numbers of early diagnoses, and the other methods are not sufficiently accurate. Id. One study reported that only 15% of patients who suffered from ovarian cancer were diagnosed with the disease at the time of their pelvic examination. Look, supra at

174. Moreover, the CA-125 test is prone to giving false positives in pre-menopausal women and has been reported to be of low predictive value in post-menopausal women. Id. at 174-75. Although transvaginal ultrasonographyis now the preferred procedure for screening for ovarian cancer, it is unable to distinguish reliably between benign and malignant tumors, and also cannot locate primary peritoneal malignancies or ovarian cancer if the ovary size is normal. Schilder et al., supra at 194-95. While genetic testing for mutations of the BRCA1, BRCA2, hMSH2, and hMLH1 genes is now available, these tests may be too costly for some patients and may also yield false negative or indeterminate results. Schilder et al., supra at 191-94.

The staging of ovarian cancer, which is accomplished through surgical exploration, is crucial in determining the course of treatment and management of the disease. AJCC Cancer Staging Handbook 187 (Irvin D. Fleming et al. eds., 5th ed. 1998); Burdette, supra at 170; Memarzadeh & Berek, supra; Shridhar et al., supra. Staging is performed by reference to the classification system developed by the International Federation of Gynecology and Obstetrics. David H. Moore, Primary Surgical Management of Early Epithelial Ovarian Carcinoma, in Ovarian Cancer 203 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001); Fleming et al. eds., supra at 188. Stage I ovarian cancer is characterized by tumor growth that is limited to the ovaries and is comprised of three substages. Id. In substage IA, tumor growth is limited to one ovary, there is no tumor on the external surface of the ovary, the ovarian capsule is intact, and no malignant cells are present in ascites or peritoneal washings. Id. Substage IB is identical to A1, except that tumor growth is limited to both ovaries. Id. Substage IC refers to the presence of tumor growth limited to one or both ovaries, and also includes one or more of the following characteristics: capsule rupture, tumor growth on the surface of one or both ovaries, and malignant cells present in ascites or peritoneal washings. Id.

Stage II ovarian cancer refers to tumor growth involving one or both ovaries, along with pelvic extension. Id. Substage IIA involves extension and/or implants on the uterus and/or fallopian tubes, with no malignant cells in the ascites or peritoneal washings, while substage IIB involves

extension into other pelvic organs and tissues, again with no malignant cells in the ascites or peritoneal washings. Id. Substage IIC involves pelvic extension as in IIA or IIB, but with malignant cells in the ascites or peritoneal washings. Id.

Stage III ovarian cancer involves tumor growth in one or both ovaries, with peritoneal metastasis beyond the pelvis confirmed by microscope and/or metastasis in the regional lymph nodes. Id. Substage IIIA is characterized by microscopic peritoneal metastasis outside the pelvis, with substage IIIB involving macroscopic peritoneal metastasis outside the pelvis 2 cm or less in greatest dimension. Id. Substage IIIC is identical to IIIB, except that the metastisis is greater than 2 cm in greatest dimesion and may include regional lymph node metastasis. Id. Lastly, Stage IV refers to the presence distant metastasis, excluding peritoneal metastasis. Id.

While surgical staging is currently the benchmark for assessing the management and treatment of ovarian cancer, it suffers from considerable drawbacks, including the invasiveness of the procedure, the potential for complications, as well as the potential for inaccuracy. Moore, supra at 206-208, 213. In view of these limitations, attention has turned to developing alternative staging methodologies through understanding differential gene expression in various stages of ovarian cancer and by obtaining various biomarkers to help better assess the progression of the disease. Vartiainen, J. et al., Int'l J. Cancer, 95(5): 313-16 (2001); Shridhar et al. supra; Baekelandt, M. et al., J. Clin. Oncol. 18(22): 3775-81.

The treatment of ovarian cancer typically involves a multiprong attack, with surgical intervention serving as the foundation of treatment. Dennis S. Chi & William J. Hoskins, Primary Surgical Management of Advanced Epithelial Ovarian Cancer, in Ovarian Cancer 241 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). For example, in the case of epithelial ovarian cancer, which accounts for ~90% of cases of ovarian cancer, treatment typically consists of: (1) cytoreductive surgery, including total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, and lymphadenectomy, followed by (2) adjuvant chemotherapy with paclitaxel and either cisplatin or carboplatin. Eltabbakh, G.H. & Awtrey, C.S., Expert Op.

Pharmacother. 2(10): 109-24. Despite a clinical response rate of 80% to the adjuvant therapy, most patients experience tumor recurrence within three years of treatment. Id. Certain patients may undergo a second cytoreductive surgery and/or second-line chemotherapy. Memarzadeh & Berek, supra.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of ovarian cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop ovarian cancer, for diagnosing ovarian cancer, for monitoring the progression of the disease, for staging the ovarian cancer, for determining whether the ovarian cancer has metastasized, and for imaging the ovarian cancer. There is also a need for better treatment of ovarian cancer.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovaries; identify and monitor ovary tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered ovary tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to ovary cells and/or ovary tissue. These ovary specific nucleic acids (OSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the OSNA is genomic DNA, then the OSNA is an ovary specific gene (OSG). In a

preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovary. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 77 through 129. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 76. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding an OSP, or that selectively hybridize or exhibit substantial sequence similarity to an OSNA, as well as allelic variants of a nucleic acid molecule encoding an OSP, and allelic variants of an OSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes an OSP or that comprises a part of a nucleic acid sequence of an OSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of an OSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of an OSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of an OSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of an OSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is an OSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of an OSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovaries. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring ovary tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered ovary tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovaries. The invention provides methods of using the polypeptides of the invention to identify and/or monitor ovary tissue, and to produce engineered ovary tissue.

The agonists and antagonists of the instant invention may be used to treat ovarian cancer and non-cancerous disease states in ovaries and to produce engineered ovary tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various

general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog,

internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3)

is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for

use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is

referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its

complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at

temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) – 0.63 (% formamide) - (600/l)

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify

nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula: $T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAS. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra.*

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding an OSP or is an OSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA

molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which

they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell

but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises an OSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A

substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and

antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or

pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$--, and --CH_2SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred

embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ε-N,N,N-trimethyllysine, ε-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the

wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which

can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be

produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.*, Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.*, Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For

instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "ovary specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the ovary as compared to other tissues in the body. In a preferred embodiment, a "ovary specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "ovary specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid

hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the ovary or to ovary cells or tissue or that are derived from such nucleic acid molecules. These isolated ovary specific nucleic acids (OSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovary, an ovary-specific polypeptide (OSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 77 through 129. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 76.

AN OSNA may be derived from a human or from another animal. In a preferred embodiment, the OSNA is derived from a human or other mammal. In a more preferred embodiment, the OSNA is derived from a human or other primate. In an even more preferred embodiment, the OSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding an OSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode an OSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes an OSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 77 through 129. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 76.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under low stringency conditions. In a more

preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding an OSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 77 through 129. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 76. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding an OSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human OSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 77 through 129. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding an OSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 77 through 129, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding an OSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding an OSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to an OSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 76. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity

with an OSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 76, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with an OSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with an OSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to an OSNA or to a nucleic acid molecule encoding an OSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the OSNA or the nucleic acid molecule encoding an OSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 77 through 129 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 76. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the OSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of an OSNA. Further, the substantially

similar nucleic acid molecule may or may not be an OSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is an OSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of an OSNA or a nucleic acid encoding an OSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes an OSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is an OSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 76. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is an OSP. However, in a preferred embodiment, the part encodes an OSP. In one aspect, the invention comprises a part of an OSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to an OSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of an OSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes an OSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as α-^{32}P-dATP, α-^{32}P-dCTP, α-^{32}P-dGTP, α-^{32}P-dTTP, α-^{32}P-3'dATP, α-^{32}P-ATP, α-^{32}P-CTP, α-^{32}P-GTP, α-^{32}P-UTP, α-^{35}S-dATP, α-^{35}S-GTP, α-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade

Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994);

Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.,* Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having

normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S.

patents that teach the preparation of PNA compounds include, but are not limited to, U.S Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly

comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of an OSNA, such as deletions, insertions, translocations, and duplications of the OSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify OSNA in, and

isolate OSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to OSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding an OSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 77 through 129. In another preferred embodiment, the probe or primer is derived from an OSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 76.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer

in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by

reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for

example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* , *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast α-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the OSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PltetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two

copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α-agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of

protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,*. Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.,* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.,* Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and

cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide OSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and

racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database http://www.abrf.org/ABRF/Research Committees/deltamass/deltamass.html (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) and http://www.glycosuite.com/ (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) and http://www.cbs.dtu.dk/databases/OGLYCBASE/ (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) and http://www.cbs.dtu.dk/databases/PhosphoBase/ (accessed October 19, 2001); or http://pir.georgetown.edu/pirwww/search/textresid.html (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another

common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method

known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational

modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be

able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from ovary are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human ovary cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See*, *e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^+ or K^+, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.),

Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is an ovary specific polypeptide (OSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 77 through 129. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of an OSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 77 through 129. A polypeptide that comprises only a fragment of an entire OSP may or may not be a polypeptide that is also an OSP. For instance, a full-length polypeptide may be ovary-specific, while a fragment thereof may be found in other tissues as well as in ovary. A polypeptide that is not an OSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-OSP antibodies. However, in a preferred embodiment, the part or fragment is an OSP. Methods of determining whether a polypeptide is an OSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of

which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, an OSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992),

supra; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably an OSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably an OSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be ovary-specific. In a preferred embodiment, the mutein is ovary-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 77 through 129. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino

acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is ovary-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to an OSP. In an even more preferred embodiment, the polypeptide is homologous to an OSP selected from the group having an amino acid sequence of SEQ ID NO: 77 through 129. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to an OSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 77 through 129. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to an OSNA. In a preferred

embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to an OSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSNA is selected from the group consisting of SEQ ID NO: 1 through 76. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes an OSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the OSP is selected from the group consisting of SEQ ID NO: 77 through 129.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 77 through 129. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the OSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of an OSP. Further, the homologous protein may or may not encode polypeptide that is an OSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is an OSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated

proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding an OSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 77 through 129. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 76.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 77 through 129, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);

Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514,

Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-OSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 77 through 129. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to an OSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$-- and --CH_2SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of an OSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.*, Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC

derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-

methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is an OSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 77 through 129, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 76, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 76.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino

acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*,

Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, α-amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine

at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the OSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize OSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly OSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of OSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of OSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.,* Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal,

polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is an OSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 77 through 129, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on an OSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of an OSP may be indicative of cancer. Differential degradation of the C or N-terminus of an OSP may also be a marker or target for anticancer therapy. For example, an OSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-OSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human ovary.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by

reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.,* Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997);

and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'$_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention

includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish

peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding an OSP. In a preferred embodiment, the OSP comprises an amino acid sequence selected from SEQ ID NO: 77 through 129, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the

transgenic cells and non-human organism comprise an OSNA of the invention, preferably an OSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 76, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human OSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)).

The present invention provides for transgenic animals that carry the transgene (*i.e.*, a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene

may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce ovaryies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a

cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.,* Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.,* Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures,

including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 76 and SEQ ID NO: 77 through 129 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Ovarian Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of an OSNA or an OSP in a human patient that has or may have ovarian cancer, or who is at risk of developing ovarian cancer, with the expression of an OSNA or an OSP in a normal human control. For purposes of the present invention, "expression of an OSNA" or "OSNA expression" means the quantity of OSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of an OSP" or "OSP expression" means the amount of OSP that can be measured by any method known in the art or the level of translation of an OSG OSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing ovarian cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of OSNA or OSP in cells, tissues, organs or bodily fluids compared with levels of OSNA or OSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of an OSNA or

OSP in the patient versus the normal human control is associated with the presence of ovarian cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in the structure of the mRNA of an OSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in an OSP compared to an OSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the OSP or subcellular OSP localization.

In a preferred embodiment, the expression of an OSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 77 through 129, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the OSNA expression that is measured is the level of expression of an OSNA mRNA selected from SEQ ID NO: 1 through 76, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. OSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. OSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of an OSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, OSNA expression may be compared to a known control, such as normal ovary nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of an OSP is measured by determining the level of an OSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 77 through 129, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of OSNA or OSP compared to normal control bodily fluids, cells, or

tissue samples may be used to diagnose the presence of ovarian cancer. The expression level of an OSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the OSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the OSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to an OSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-OSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the OSP will bind to the anti-OSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-OSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the OSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an OSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure OSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-OSP antibody is attached to a solid support and an allocated amount of a labeled OSP and a sample of interest are incubated with the solid support. The amount of labeled OSP detected which is attached to the solid support can be correlated to the quantity of an OSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential

separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of an OSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more OSNAs of interest. In this approach, all or a portion of one or more OSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen

tested for expression of OSNA or OSP includes, without limitation, ovary tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, ovary cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary ovarian cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and breast. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in OSNAs or OSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of an OSNA or OSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other OSNA or OSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular OSNA or OSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having ovarian cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP and then ascertaining whether the patient has ovarian cancer from the expression level of the OSNA or OSP. In general, if high expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at

least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether ovarian cancer has metastasized in a patient. One may identify whether the ovarian cancer has metastasized by measuring the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a variety of tissues. The presence of an OSNA or OSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of an OSNA or OSP is associated with ovarian cancer. Similarly, the presence of an OSNA or OSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of an OSNA or OSP is associated with ovarian cancer. Further, the presence of a structurally altered OSNA or OSP that is associated with ovarian cancer is also indicative of metastasis.

In general, if high expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The OSNA or OSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with ovarian cancers or other ovary related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of ovarian disorders.

Staging

The invention also provides a method of staging ovarian cancer in a human patient. The method comprises identifying a human patient having ovarian cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more OSNAs or OSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more OSNAs or OSPs is determined for each stage to obtain a standard expression level for each OSNA and OSP. Then, the OSNA or OSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The OSNA or OSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the OSNAs and OSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of an OSNA or OSP to determine the stage of an ovarian cancer.

Monitoring

Further provided is a method of monitoring ovarian cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the ovarian cancer. The method comprises identifying a human patient that one wants to monitor for ovarian cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more OSNAs or OSPs, and comparing the OSNA or OSP levels over time to those OSNA or OSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in an OSNA or OSP that are associated with ovarian cancer.

If increased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or

failure to progress to a neoplastic lesion. If decreased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of OSNAs or OSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of ovarian cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of an OSNA and/or OSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more OSNAs and/or OSPs are detected. The presence of higher (or lower) OSNA or OSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly ovarian cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more OSNAs and/or OSPs of the invention can also be monitored by analyzing levels of expression of the OSNAs and/or OSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in an OSG, thereby determining if a human with the genetic lesion is susceptible to developing ovarian cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing ovarian cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the OSGs of this invention, a chromosomal rearrangement of OSG, an aberrant modification of OSG (such as of the methylation pattern of the genomic DNA), or allelic loss of an OSG. Methods to detect such lesions in the OSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Ovarian Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a patient suspected of having or known to have a noncancerous ovarian disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP, comparing the expression level or structural alteration of the OSNA or OSP to a normal ovary control, and then ascertaining whether the patient has a noncancerous ovarian disease. In general, if high expression relative to a control of an OSNA or OSP is indicative of a particular noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of a noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether an OSNA and/or OSP is associated with a particular noncancerous ovarian disease by obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining which OSNAs and/or OSPs are expressed in the tissue at either a higher or a lower level than in normal ovary tissue. In another embodiment, one may determine whether an OSNA or OSP exhibits structural alterations in a particular noncancerous ovarian disease state by obtaining ovary tissue from a patient having a noncancerous ovarian disease of interest and determining the structural alterations in one or more OSNAs and/or OSPs relative to normal ovary tissue.

Methods for Identifying Ovary Tissue

In another aspect, the invention provides methods for identifying ovary tissue. These methods are particularly useful in, *e.g.*, forensic science, ovary cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is ovary tissue or has ovary tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising ovary tissue or having ovary tissue-like characteristics, determining whether the sample expresses one or more OSNAs and/or OSPs, and, if the sample expresses one or more OSNAs and/or OSPs, concluding that the sample comprises ovary tissue. In a preferred embodiment, the OSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 77 through 129, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 76, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses an OSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether an OSP is expressed. Determining whether a sample expresses an OSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the OSP has an amino acid sequence selected from SEQ ID NO: 77 through 129, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two OSNAs and/or OSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five OSNAs and/or OSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is ovary tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into ovary tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new ovary tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Ovary Tissue

In another aspect, the invention provides methods for producing engineered ovary tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing an OSNA or an OSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of ovary tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal ovary tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered ovary tissue or cells comprises one of these cell types. In another embodiment, the engineered ovary tissue or cells comprises more than one ovary cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the ovary cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more OSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode OSPs having amino acid sequences selected from SEQ ID NO: 77 through 129, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 76, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, an OSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial ovary tissue may be used to treat patients who have lost some or all of their ovary function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises an OSNA or part thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 76, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises an OSP or fragment thereof. In a more preferred embodiment, the OSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 77 through 129,

a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-OSP antibody, preferably an antibody that specifically binds to an OSP having an amino acid that is selected from the group consisting of SEQ ID NO: 77 through 129, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example OSP polypeptide, fusion protein, or fragments thereof, antibodies specific for OSP, agonists, antagonists or inhibitors of OSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic

agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention , *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of ovary function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) , for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is

preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of an OSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of an OSP are administered, for example, to complement a deficiency in the native OSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes an OSP having the amino acid sequence of SEQ ID NO: 77 through 129, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express an OSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in OSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode an OSP having the amino acid sequence of SEQ ID NO: 77 through 129, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of an OSG antisense nucleic acid, are administered to downregulate transcription and/or translation of an OSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of an OSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to OSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and

Sigurdsson *et al., Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the OSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.,* Intody *et al., Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al., Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding an OSP, preferably an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 76, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an OSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant OSP defect.

Protein compositions are administered, for example, to complement a deficiency in native OSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to OSP. The immune response can be used to modulate activity of OSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate OSP.

In a preferred embodiment, the polypeptide is an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1

through 76, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of OSP, or to target therapeutic agents to sites of OSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 76, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to an OSP or have a modulatory effect on the expression or activity of an OSP. Modulators which decrease the expression or activity of OSP (antagonists) are believed to be useful in treating ovarian cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of an OSP can also be designed, synthesized and tested for use in the imaging and treatment of ovarian cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the OSPs identified herein. Molecules identified in the library as being capable of binding to an OSP are key candidates for further evaluation for use in the treatment of ovarian cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of an OSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of OSP is administered. Antagonists of OSP can be produced using methods generally known in the art. In particular, purified OSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of an OSP.

In other embodiments a pharmaceutical composition comprising an agonist of an OSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP comprising an amino acid sequence of SEQ ID NO: 77 through 129, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 76, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Ovary Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the ovary or to specific cells in the ovary. In a preferred embodiment, an anti-OSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if ovary tissue needs to be selectively destroyed. This would be useful for targeting and killing ovarian cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting ovary cell function.

In another embodiment, an anti-OSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring ovary function, identifying ovarian cancer tumors, and identifying noncancerous ovarian diseases.

EXAMPLES

Example 1: Gene Expression analysis

OSGs were identified by mRNA subtraction analysis using standard methods. The sequences were extended using GeneBank sequences, Incyte's proprietary database. From the nucleotide sequences, predicted amino acid sequences were prepared. DEX0310_1, DEX0310_2 correspond to SEQ ID NO:1, 2 etc. DEX0161 and DEX0168 were the parent sequences found in the mRNA subtractions. The sequences listed as flexDEX are sequences prepared by in silico sequence extension. The sequences beginning with DEX0310_77 are the predicted amino acid sequences.

DEX0310_1 DEX0161_1 DEX0310_77

DEX0310_1	DEX0161_1	DEX0310_77
DEX0310_2	DEX0161_2	DEX0310_78
DEX0310_3	DEX0161_3	
DEX0310_4	DEX0161_4	DEX0310_79
DEX0310_5	DEX0161_5	DEX0310_80
DEX0310_6	DEX0161_6	
DEX0310_7	DEX0161_7	DEX0310_81
DEX0310_8	flex DEX0161_7	DEX0310_82
DEX0310_9	DEX0161_8	
DEX0310_10	flex DEX0161_8	DEX0310_83
DEX0310_11	DEX0161_9	DEX0310_84
DEX0310_12	DEX0161_10	DEX0310_85
DEX0310_13	DEX0161_11	DEX0310_86
DEX0310_14	DEX0161_12	DEX0310_88
DEX0310_15	DEX0161_13	DEX0310_89
DEX0310_16	DEX0161_14	DEX0310_90
DEX0310_17	DEX0161_15	
DEX0310_18	DEX0161_16	DEX0310_91
DEX0310_19	DEX0161_17	DEX0310_92
DEX0310_20	DEX0161_18	DEX0310_93
DEX0310_21	DEX0161_19	DEX0310_94
DEX0310_22	flex DEX0161_19	
DEX0310_23	DEX0161_20	DEX0310_95
DEX0310_24	DEX0168_1	DEX0310_96
DEX0310_25	flex DEX0168_1	
DEX0310_26	DEX0168_3	DEX0310_97
DEX0310_27	DEX0168_4	DEX0310_98
DEX0310_28	flex DEX0168_4	
DEX0310_29	DEX0168_5	DEX0310_99
DEX0310_30	flex DEX0168_5	
DEX0310_31	DEX0168_6	
DEX0310_32	flex DEX0168_6	
DEX0310_33	DEX0168_7	DEX0310_100
DEX0310_34	flex DEX0168_7	
DEX0310_35	DEX0168_8	DEX0310_101
DEX0310_36	flex DEX0168_8	
DEX0310_37	DEX0168_9	DEX0310_102
DEX0310_38	flex DEX0168_9	DEX0310_103
DEX0310_39	DEX0168_10	DEX0310_104
DEX0310_40	flex DEX0168_10	DEX0310_105
DEX0310_41	DEX0168_11	DEX0310_106
DEX0310_42	flex DEX0168_11	
DEX0310_43	DEX0168_12	DEX0310_107
DEX0310_44	DEX0168_13	DEX0310_108
DEX0310_45	flex DEX0168_13	DEX0310_109
DEX0310_46	DEX0168_14	
DEX0310_47	DEX0168_15	DEX0310_110
DEX0310_48	DEX0168_16	DEX0310_111
DEX0310_49	flex DEX0168_16	DEX0310_112
DEX0310_50	DEX0168_17	DEX0310_113
DEX0310_51	flex DEX0168_17	
DEX0310_52	DEX0168_18	DEX0310_114
DEX0310_53	flex DEX0168_18	
DEX0310_54	DEX0168_19	DEX0310_115
DEX0310_55	flex DEX0168_19	
DEX0310_56	DEX0168_20	DEX0310_116
DEX0310_57	flex DEX0168_20	
DEX0310_58	DEX0168_21	DEX0310_117
DEX0310_59	flex DEX0168_21	
DEX0310_60	DEX0168_22	DEX0310_118
DEX0310_61	DEX0168_23	DEX0310_119
DEX0310_62	flex DEX0168_23	

DEX0310_63	DEX0168_24	DEX0310_120
DEX0310_64	DEX0168_25	DEX0310_121
DEX0310_65	flex DEX0168_25	DEX0310_122
DEX0310_66	DEX0168_26	DEX0310_123
DEX0310_67	flex DEX0168_26	
DEX0310_68	DEX0168_27	DEX0310_124
DEX0310_69	flex DEX0168_27	
DEX0310_70	DEX0168_28	DEX0310_125
DEX0310_71	flex DEX0168_28	DEX0310_126
DEX0310_72	DEX0168_29	DEX0310_127
DEX0310_73	DEX0168_30	DEX0310_128
DEX0310_74	flex DEX0168_30	
DEX0310_75	DEX0168_31	DEX0310_129
DEX0310_76	flex DEX0168_31	

Example 1a: ATCC Deposit Information

The table below summarizes the information corresponding to each OSG depicted in provisional application Serial No. 60/268,834, filed February 15, 2001, which is herein incorporated by reference in its entirety and which is referred to as DEX0161.

The cDNAs of the OSGs were deposited on the date listed in the column entitled ATCC Deposit Date. Each cDNA was cloned with vector PCR2.1 (Invitrogen, San Diego, CA). The "Contig Length" is the number of nucleotides in the contig identified by Contig ID and DEX0161 ID #. The "CloneSeq Length" is the number of nucleotides in the clone with "Clone ID" number and deposited with the ATCC.

The deposited material in the sample assigned ATCC Deposit Number in the table for any cDNA clone also contains one or more additional plasmids, each having a cDNA different from a given clone. Thus, deposits sharing the same ATCC number contain at least a plasmid for each "Clone ID" identified in the table. Typically, each ATCC deposit contains a mixture of approximately equal amounts by weight of about fifty plasmids, each containing a different cDNA clone.

The bioassays used were:

Psovr003: 3 cancer-papillary carcinoma grade2+papillary serous and endometrioid carcinoma grade 3+papillary serous adenocarcinoma grad 2 substracted with a mixture of normal tissues - kidney+pancreas+ spleen+small intestine+heart+colon

Psovr005: one matching sample-papillary serous carcinoma grade 3

Psovr007: 3 cancer -papillary carcinoma grade 2+papillary serous and endometrioid carcinoma grade 3+papillary serous adenocarcinoma grade 2 substracted with a mixture of five normal ovaries

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in the Table below. First, a plasmid is

directly isolated by screening the clones using a polynucleotide probe corresponding to clone id, e.g., 601038725F1.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance with ^{33}P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g. Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated in 1.5% agar plates (containing the appropriate selection agent, e.g. ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two printers of 17-20 nucleotides derived from both ends of the DEX0161 ID NO:X (i.e., within the region of DEX0161 ID NO:X bounded by the 5' NT and the 3'NT of the clone defined in the table below) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 µl of reaction mixture with 0.5 µg of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 minute; annealing at 55°C for 1 minute; elongation at 72°C for 1 minute) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar or identical to 5' RACE is available for generating the

missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7); 1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

DEX0161 ID NO:	Contig Clone Id	Contig length	Clone ID	Clone Length	ATCC No and Date	Bioassay
1	23.104	355	601038725F1	355	Feb. 15, 2001 PTA3064	PSovr007
2	23.108	957	601630993F1	957	Feb. 15, 2001 PTA3064	PSovr005
3	23.16	738	601628602F1	738	Feb. 15, 2001 PTA3064	PSovr005
4	23.167	586	601632286F1	586	Feb. 15, 2001 PTA3064	PSovr007
	23.167		601633105F1	785	Feb. 15, 2001 PTA3064	PSovr007
	23.167		601633138F1	733	Feb. 15, 2001 PTA3064	PSovr007
	23.167		601633320F1	841	Feb. 15, 2001 PTA3064	PSovr007
	23.167		601633338F1	801	Feb. 15, 2001 PTA3064	PSovr007
	23.167		601633442F1	800	Feb. 15, 2001 PTA3064	PSovr007
	23.167		601633455F1	443	Feb. 15, 2001 PTA3064	PSovr007
5	23.175	772	601630092F1	772	Feb. 15, 2001 PTA3064	PSovr005

	23.175		601630331F1	771	Feb. 15, 2001 PSovr005 PTA3064
6	23.186	686	601628695F1	686	Feb. 15, 2001 PSovr005 PTA3064
7	23.193	720	601629479F1	720	Feb. 15, 2001 PSovr005 PTA3064
	23.193		601629822F1	692	Feb. 15, 2001 PSovr005 PTA3064
	23.193		601629847F1	672	Feb. 15, 2001 PSovr005 PTA3064
8	23.2	1878	600981916F1	1878	Feb. 15, 2001 PSovr005 PTA3064
9	23.203	666	601629424F1	666	Feb. 15, 2001 PSovr005 PTA3064
10	23.204	194	000000092D1	194	Feb. 15, 2001 PSovr003 PTA3064
	23.204		600973748F1	1001	Feb. 15, 2001 PSovr003 PTA3064
	23.204		600975714F1	1804	Feb. 15, 2001 PSovr003 PTA3064
	23.204		600976861F1	570	Feb. 15, 2001 PSovr003 PTA3064
11	23.35	600	601625756F1	600	Feb. 15, 2001 PSovr003 PTA3064
12	23.377	829	601625763F1	885	Feb. 15, 2001 PSovr003 PTA3064
13	23.56	906	601633634F1	906	Feb. 15, 2001 PSovr007 PTA3064
14	23.68	744	601634123F1	744	Feb. 15, 2001 PSovr007 PTA3064
15	23.74	925	601633501F1	925	Feb. 15, 2001 PSovr007 PTA3064
16	23.781	766	601634180F1	737	Feb. 15, 2001 PSovr007 PTA3064
17	23.829	133	601633581F1	774	Feb. 15, 2001 PSovr007 PTA3064
18	23.894	280	601630616F1	753	Feb. 15, 2001 PSovr005 PTA3064
	23.894		601630888F1	692	Feb. 15, 2001 PSovr005 PTA3064
	23.894		601630949F1	760	Feb. 15, 2001 PSovr005 PTA3064
	23.894		601630967F1	633	Feb. 15, 2001 PSovr005 PTA3064
19	23.94	926	601634034F1	900	Feb. 15, 2001 PSovr007 PTA3064
20	23.982	1247	601631118F1	804	Feb. 15, 2001 PSovr007 PTA3064

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse

Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the OSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the OSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the OSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 76 being a diagnostic marker for cancer.

Example 3: Protein Expression

The OSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the OSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the OSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of OSNA, and six histidines, flanking the COOH-terminus of the coding sequence of OSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of OSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. OSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the

vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
DEX0310_100			Ck2_Phospho_Site 4-7; Myristyl 17-22;23-28;	
DEX0310_101		1,o27-49i	Myristyl 6-11;7-12;	
DEX0310_102		2,i21-43o53-75i	Asn_Glycosylation 69-72; Ck2_Phospho_Site 42-45; Tyr_Phospho_Site 23-30;	
DEX0310_103			Myristyl 29-34;	
DEX0310_104	45-67,1.11,23		Asn_Glycosylation 2-5; Ck2_Phospho_Site 9-12;35-38; Myristyl 14-19;25-30;53-58; Pkc_Phospho_Site 6-8;19-21;57-59;95-97;	
DEX0310_105	260-271,1.26,12 319-328,1.23,10 150-178,1.06,29		Asn_Glycosylation 435-438;704-707; Camp_Phospho_Site 241-244;326-329; Ck2_Phospho_Site 19-22;52-55;65-68;122-125;224-227;517-520;591-594;600-603;655-658;718-721;742-745; Gram_Pos_Anchoring 653-658; Myristyl 26-31;27-32;173-178;255-260;330-335;436-441;473-478;486-491;496-501;613-618; Pkc_Phospho_Site 52-54;82-84;220-222;355-357;373-375;474-476;514-516;627-629;726-728;735-737;757-759; Tyr_Phospho_Site 215-221;	
DEX0310_106	29-42,1.02,14		Pkc_Phospho_Site 37-39;	
DEX0310_107		1,i7-29o	Ck2_Phospho_Site 56-59; Myristyl 57-62;	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			Pkc_Phospho_Site 4-6;33-35;	
DEX0310_108			Amidation 49-52; Ck2_Phospho_Site 6-9; Pkc_Phospho_Site 25-27;45-47;	
DEX0310_109	2-11,1.01,10		Ck2_Phospho_Site 15-18; Myristyl 21-26;	
DEX0310_110	6-50,1.23,45		Amidation 36-39;132-135; Asn_Glycosylation 12-15; Camp_Phospho_Site 70-73; Ck2_Phospho_Site 52-55;57-60;84-87;113-116; Myristyl 11-16;78-83;110-115; Pkc_Phospho_Site 42-44;57-59;73-75;81-83;	
DEX0310_111			Ck2_Phospho_Site 40-43; Glycosaminoglycan 60-63; Myristyl 20-25;61-66;94-99; Tyr_Phospho_Site 33-39;	
DEX0310_112			Ck2_Phospho_Site 40-43;100-103; Myristyl 11-16;	
DEX0310_113			Ck2_Phospho_Site 16-19;33-36;	
DEX0310_114			Ck2_Phospho_Site 13-16; Pkc_Phospho_Site 7-9;12-14;13-15;	
DEX0310_115	55-75,1.14,21		Amidation 111-114; Asn_Glycosylation 71-74; Camp_Phospho_Site 114-117; Ck2_Phospho_Site 30-33;69-72;91-94; Pkc_Phospho_Site 63-65;	
DEX0310_116			Ck2_Phospho_Site 6-9; Glycosaminoglycan 68-71; Myristyl	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			19-24;66-71;67-72;69-74; Pkc_Phospho_Site 11-13;34-36;	
DEX0310_117			Asn_Glycosylation 16-19; Ck2_Phospho_Site 5-8; Myristyl 22-27;24-29;27-32; Pkc_Phospho_Site 5-7;65-67;	
DEX0310_119			Camp_Phospho_Site 6-9; Pkc_Phospho_Site 4-6;	
DEX0310_120	13-37,1.02,25		Myristyl 4-9;26-31; Pkc_Phospho_Site 13-15;14-16;	
DEX0310_121	55-65,1.15,11	1,o123-145i	Myristyl 47-52;	
DEX0310_122	9-55,1.03,47	1,i93-115o	Myristyl 40-45;54-59;	
DEX0310_123	36-46,1.13,11		Ck2_Phospho_Site 105-108;114-117;	
DEX0310_124			Camp_Phospho_Site 63-66; Pkc_Phospho_Site 61-63;62-64;	23,.941,.644
DEX0310_125	75-89,1.23,15 40-54,1.08,15		Camp_Phospho_Site 50-53; Ck2_Phospho_Site 3-6; Myristyl 36-41;40-45;	
DEX0310_126	70-83,1.18,14		Glycosaminoglycan 23-26; Myristyl 59-64;83-88; Pkc_Phospho_Site 69-71;88-90;	
DEX0310_127			Camp_Phospho_Site 22-25; Pkc_Phospho_Site 9-11;	
DEX0310_128			Camp_Phospho_Site 27-30; Myristyl 32-37;	
DEX0310_129	77-86,1.26,10		Ck2_Phospho_Site 44-47; Myristyl 83-88; Pkc_Phospho_Site 9-11;36-38;104-106;	
DEX0310_77	67-128,1.16,62		Camp_Phospho_Site 150-153;158-161;172-175; Ck2_Phospho_Site	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			38-41;95-98;144-147;175-178;204-207;215-218;238-241; Myristyl 61-66;68-73;93-98;129-134; Pkc_Phospho_Site 14-16;107-109;157-159;168-170;175-177;188-190;	
DEX0310_78			Asn_Glycosylation 32-35; Camp_Phospho_Site 19-22; Ck2_Phospho_Site 11-14;67-70;84-87; Myristyl 42-47;54-59; Pkc_Phospho_Site 22-24;27-29;67-69; Rgd 85-87;	
DEX0310_79		1,i21-43o	Tyr_Phospho_Site 14-22;	
DEX0310_80			Ck2_Phospho_Site 29-32;65-68; Myristyl 6-11;	
DEX0310_81			Amidation 34-37;42-45; Ck2_Phospho_Site 108-111; Myristyl 29-34;55-60; Pkc_Phospho_Site 11-13;	
DEX0310_82	181-196,1.14,16 71-91,1.06,21 99-116,1.02,18		Ck2_Phospho_Site 143-146;222-225;223-226; Ig_Mhc 213-219; Myristyl 33-38;84-89;88-93;122-127;129-134;180-185; Pkc_Phospho_Site 72-74;79-81;183-185;209-211; Prokar_Lipoprotein 30-40;	18,.984,.909
DEX0310_83			Myristyl 96-101;154-159; Prokar_Lipoprotein 28-38;	
DEX0310_84			Asn_Glycosylation 15-18;45-48;	
DEX0310_85	63-77,1.1,15 47-57,1,11		Ck2_Phospho_Site 37-40;68-71;	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
DEX0310_86	24-33,1.29,10		Ck2_Phospho_Site 14-17; Tyr_Phospho_Site 24-32;	
DEX0310_87	34-57,1.21,24		Myristyl 25-30;45-50; Pkc_Phospho_Site 69-71;	
DEX0310_88			Asn_Glycosylation 44-47; Ck2_Phospho_Site 15-18;35-38; Myristyl 12-17;16-21;26-31;78-83;82-87;83-88; Pkc_Phospho_Site 46-48; Tyr_Phospho_Site 48-54;	
DEX0310_89			Asn_Glycosylation 64-67;69-72; Myristyl 89-94;	
DEX0310_90		2,i7-26o63-85i	Ck2_Phospho_Site 47-50; Myristyl 17-22; Pkc_Phospho_Site 8-10;47-49;59-61; Prokar_Lipoprotein 68-78;	21,.997,.957
DEX0310_91			Ck2_Phospho_Site 25-28; Myristyl 31-36;	
DEX0310_92	34-49,1.06,16		Ck2_Phospho_Site 29-32; Myristyl 13-18;52-57;	
DEX0310_93	6-24,1.01,19		Ck2_Phospho_Site 12-15; Rgd 9-11;	
DEX0310_94	45-79,1.05,35		Ck2_Phospho_Site 93-96; Glycosaminoglycan 9-12; Myristyl 27-32;130-135; Pkc_Phospho_Site 32-34;40-42;58-60;70-72;134-136;	
DEX0310_95			Asn_Glycosylation 44-47;	23,.883,.635
DEX0310_96	34-62,1.03,29		Ck2_Phospho_Site 52-55; Pkc_Phospho_Site 6-8;43-45;54-56;	
DEX0310_97			Ck2_Phospho_Site 67-70; Pkc_Phospho_Site 34-36;	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
DEX0310_98			Camp_Phospho_Site 30-33;31-34;	
DEX0310_99			Camp_Phospho_Site 55-58; Myristyl 49-54;63-68;	

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 76. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al., Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ)

and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 μg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of

administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and

EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 μg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf

serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and

more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon,"Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of

interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or

transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 77 through 129;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 76;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.

6. A method for determining the presence of an ovary specific nucleic acid (OSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to an ovary specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to an OSNA in the sample, wherein the detection of the hybridization indicates the presence of an OSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 77 through 129; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 76.

12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.

13. A method for determining the presence of an ovary specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the ovary specific protein; and

(b) detecting binding of the antibody to an ovary specific protein in the sample, wherein the detection of binding indicates the presence of an ovary specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of ovarian cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the ovary specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the

polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of ovarian cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient.

16. A method of treating a patient with ovarian cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the ovarian cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> diaDexus, Inc.
Salceda, Susana
Macina, Roberto
Hu, Ping
Recipon, Herve
Karra, Kalpana
Cafferkey, Robert
Sun, Yongming
Liu, Chenghua

<120> Compositions and Methods Relating to Ovarian Specific Genes and Proteins

<130> DEX-0310

<150> 60/268,290
<151> 2001-02-13

<150> 60/268,834
<151> 2001-02-15

<160> 129

<170> PatentIn version 3.1

<210> 1
<211> 829
<212> DNA
<213> Homo sapien

<400> 1
cgtggtcgcg gcgaggtaca agctttttt ttttttttt ttttttttt ggcaaaaaaa 60

aataggctcc gttttatttt cttccattgg atccaatggg catcttttg aaagcctgtc 120

tgtgtgccaa cccttccccc aaaggaggtt atctcaggtg ggtggagcca agttctcacg 180

gggtggaaag gagaccgtgg acacacacaa gagaagaacc tccaaagccc tcctccatta 240

tgtggcagag aattcaacgc tgggcgtacc tcagtggctc aatagcgtgt ctccgtggtg 300

ctgacaattg tcgtacctcc gcctcacaat tctcccacca aacaaaaata tgtgacacaa 360

acacgcagcc aggggcgagc ccgaccgacg cccgcaaggc tcggcgccca aaatcaccgc 420

gaccccgacc agcgccagcg ccccgacaag cacccgggca acacccccac agcacgaccg 480

gcgcggccat cacaacgggc cccaccgccc aaagacgaga ggccaccgac gcagagaaca 540

agaggaagag aacgagacag agaacacgaa gaaccacagg gcaaacgtac gagcaaacaa 600

aaaaaagaaa aaaaaaaacc aagagggacg caggagacga cggacgcgcc agaaagacaa 660

agagacaagc aaagagaaat aagggaaaag caaaaagagg aaggtccaag caggagagaa 720

aaaagaagca acgagccaca aaacaggagc ataaagaaaa ggacagaaaa gcaccaagag 780

gacaaacaaa ggagggagaa cagaacacga aagacgaaag agaagaaaa 829

<210> 2
<211> 766
<212> DNA
<213> Homo sapien

<400> 2
atatgactca tatagcgaat ggtgcatcta atcatctcga gcggcgcagt gtgatggatc 60

gcccgggcag gtcggctact ggaggaggct gtgaggaaga aggggtcgga ggagaggagg 120

agaccccaca agggaggagg aggaagaggc tgctacccgc ccgctgccgc ccccgtactc 180

ggagacccga ctcgcatggc gtcccgatga tgcgcaggct caagaagaac ttgtgccctg 240

atgtcgggaa atgtttagtc gtactgatct gaacgtttat ctaacgtaca cttttgtatt 300

tttttttta atttgaagga acactgatga agcccgtgcc ataccctcc cgagtctaat 360

aaaacggtat aatcaaaaaa aaaaaaaaaa aaacaaaaaa aggcttgggg ggtacctctg 420

tgggccaaag gcgtggtccc gtggggtgag aaatgggtta cccggcctca aaattctccc 480

caacaacttt agagaggcaa caaccccgca aacaaacaca gagagcaagc agccagagac 540

agggcaacaa cacaaaagca cacagacaga aggagggcgc agcagagggg acacaagcga 600

cgccagagag aggccagaca caggcgcacc agaagagaag agagagaacg accgggcggg 660

aagagcagaa ggagaagcga cagcagcgag aagagcaagg gacaagaggc gacagagaga 720

ggaagaggca accggcaaac gcacaggacg ggcaacaggc gcgaaa 766

<210> 3
<211> 133
<212> DNA
<213> Homo sapien

<400> 3
acttttaaa attaaactgg taaacagaca ttagaattac caatgtaaac tttatatcta 60

aaaccaaaat caattacaaa ggatcattat tttccaaagt agatactttt tttagtgctg 120

ttcagttttt ctg 133

<210> 4
<211> 280
<212> DNA
<213> Homo sapien

<400> 4
cggccgtccg ggcaggtatt taattgatta gaagacaagt tttaaggtat tttataatga 60

cttccttatg acaaatatct catcagaaat ctactgccta tctgatatct gactccatgc 120

gtgtatacgc gtgtagctca gtctattccc agcatagagg gtcatttgat gtacacgtct 180

atttgtatta ccatggctac gttggtgtca ctaccttgac gatgatattt agcagtgtac 240

tttttgggta tggttttggg gttatttggt tgttgttaaa 280

<210> 5
<211> 1247
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (347)..(347)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (446)..(446)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (514)..(515)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (666)..(666)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (732)..(732)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (382)..(382)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (981)..(982)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (997)..(998)
<223> a, c, g or t

<400> 5
aaaaaaaaaa aaaagaagaa tgaatacgac acactatagg gtccttggtt tctctagatg 60

actgctcgag ccgcgactta tgtgatggat gagcgagacg cccgggcagg tacctggaga 120

caggtgcagt ccctcacctg tgaaggtgga tgcccttgaa atggaaccaa tgagtccctg 180

ggaacgccta gactgcgtga aactgcgaag cagagacgtt ggccgctctg cacacgctgc 240

ctacatcgta ccatgcacac atatctgcgc ccgcctggcg agcgatggag acttccacga 300

actcattgaa ggcacgtgaa gactgtgctc tccgtcatcc tgttcgntgg aaatagccaa 360

cgcaacctga cctcgaatct gnttagagta atggtgcccg tagcagcatc ttggagtact 420

ctggtcgcac cgtagcacgg tcaagncttg tctggcaagt gtccgagtgg gaggagcgag 480

gaatattcgt ccttgctctg ctctcctata tacnncttgt tcctaccact cgagaaccgt 540

ggaggggggat agagcgtaaa ctcgttggac tctcatcgtc ccccatctag gcttagaagt 600

acctcatcgt gcgaagagac atgggacgac cgtgtgctgg gaatttaccc gggccggctc 660

tcgtanagaa catgatgaag gaatgagatc taggggatta actccaagac cgtagagagc 720

tatcctctaa gnctagtaca accgcattcg gatccttttg cacataaaac acattggcaa 780

caacaccctc agtggcaagc caagtgagta ttaacccagg cgagagcaaa taacagagaa 840

gcacacatcc cgataaatct taagagcaaa accagtccct ttccaagttg ggcctttctg 900

gccagggttt tccctccagg gaagcggcaa ggactagcat tctcgggaca ctacgcgaca 960

taaaggcctc ttagcttcct nntaaacaac ccgaacnnac tccttaccaa ggacttctac 1020

gaacacacac aggtttacca aacataaatc taggtttacc aggaaaaaat ctttgtgcct 1080

agaaataact ttttacacag ggtatttttg aataccatct aaaactggct ttttttttcc 1140

ccagcaagta tcccaccaac tgtggtcctg gcttcaataa atcttgggga aaactccgaa 1200

aaaaacaaaa accaaaacaa aaggcggggg aacaggggca caccggc 1247

<210> 6
<211> 355
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (306)..(306)
<223> a, c, g or t

<400> 6
gcccgatgaa gatcaccaca tgggcgatgg tgccctagat gcatgccgag cggcgcagtg 60

tgatggatac gaatgtaaat attggtgtcg ttgctcgaca ttttagactt gaaagcgata 120

tgctgcgagt ataatgtagt taaccatatt aggtggcgag atattcaata aatagtttac 180

atctgtcgaa aaaaaaaaaa aaaaaaaaaa aaaaggcgcg ggggggtccc ccggggccca 240

aggcggggcc ccgggggggaa attggttccc gcggccccaa attccccccа aaaataagaa 300

aaaacntggg acaaccaact ccccccgacct cccctccaac ccacccacaa caaaa 355

<210> 7
<211> 957
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (519)..(519)
<223> a, c, g or t

<400> 7
ttgggcttta ctaatgcatg ctcgagcggc cgccaagtgt gatggatgcg tggtcgcggc 60

cgaggtacgc ggtcagcagc tatctgagcc tgacggcctg atgcagtgtg agagtcccac 120

agaagctaca gcctgcacag agtcttagct gcatgaaggg agacaccgtg gatgaagaca 180

gtggtcccct acagaaatgt tctataggtt ctctaacacg ctcagccccc actaccaatg 240

gcgagactag cacgctgcag ggatcccaag ggagaggggt ctctccatcc acaccacaca 300

agggcgagtc aaagccctta tcatcgcgca tgtcgacgtc atgtaaaagc gcctacaaat 360

aagatattct gcacttggtt gaaatgtcct acatacataa caaagacaca tactcaacta 420

cacggagtcg atcgatcacc ggtccgtgcg ggcgaatgcc acttcgctct cgtgcgtcca 480

atgatgactc atagttacac accggtgtgc ggcgcacanc tatgagtgga tattcgcccg 540

ggatcacaga ccatggattt ccccccgtgg ctgatcgaat atgcggtacg cggcatcaaa 600

ttcgcccggg agctacagac ctaaaaaagt tgaccgcgca gcggccgaag aacaggcttt 660

cgacggaatg ccaaacacag agggccgcag accggcaggc gaccccgggc ggaggagccc 720

cactgcggca gggcgaggcg aaggacagat acgaggacgc gagccacacg cgcgcccgtg 780

catgagacgg agacggccga gggagcgcag acccgaagca gcgcgccaag agcgaccgcg 840

caagccacac gcgcctaggc cctgcgccac ggccggccac gcgcgagagg cggggcggag 900

caccgcagga gaccgaccac ggacccgacc ggcccagggc agcagagcca ccgagct 957

<210> 8
<211> 1460
<212> DNA

<213> Homo sapien

<220>
<221> misc_feature
<222> (1022)..(1022)
<223> a, c, g or t

<400> 8
ggcctgggct ctgctattcc tcaccctcct cactcagggc acagggtcct gggcccagtc 60

tgccctgact cagtctgcct ccgtgtctgg gtctcctgga cagtcgatca ccatctcctg 120

cactggaacc agcagtcacg ttggtggtta taactatgtc tcctggtacc aacagcaccc 180

aggcaaagcc cccaaactca tcatttatga ggtcagtaat cggccctcag gggtttctaa 240

tcgcttctct ggctccaagt ctggcaacac ggcctccctg accatctctg ggctccaggc 300

tgaggacgag gctgattatt actgctgctc atatacaaga agtacttctc atgtcttcgg 360

aactgggacc aaggtcaccg tcctaggtca gcccaaggcc aaccccactg tcactctgtt 420

cccgccctcc tctgaggagc tccaagccaa caaggccaca ctagtgtgtc tgatcagtga 480

cttctacccg ggagctgtga cagtggcctg gaaggcagat ggcagccccg tcaaggcggg 540

agtggagacc accaaaccct ccaaacagag caacaacaag tacgcggcca gcagctacct 600

gagcctgacg cccgagcagt ggaagtccca cagaagctac agctgccagg tcacgcatga 660

agggagcacc gtggatgaag acagtggtcc cctacagaaa tgttctatag gttctctaac 720

acgctcagcc cccactacca atggcgagac tagcacgctg cagggatccc aagggagagg 780

ggtctctcca tccacaccac acaagggcga gtcaaagccc ttatcatcgc gcatgtcgac 840

gtcatgtaaa agcgcctaca ataagatat tctgcacttg gttgaaatgt cctacataca 900

taacaaagac acatactcaa ctacacggag tcgatcgatc accggtccgt gcgggcgaat 960

gccacttcgc tctcgtgcgt ccaatgatga ctcatagtta cacaccggtg tgcggcgcac 1020

anctatgagt ggatattcgc ccgggatcac agaccatgga tttccccccg tggctgatcg 1080

aatatgcggt acgcggcatc aaattcgccc gggagctaca gacctaaaaa agttgaccgc 1140

gcagcggccg aagaacaggc tttcgacgga atgccaaaca cagagggccg cagaccggca 1200

ggcgaccccg ggcggaggag ccccactgcg gcagggcgag gcgaaggaca gatacgagga 1260

cgcgagccac acgcgcgccc gtgcatgaga cggagacggc cgagggagcg cagacccgaa 1320

gcagcgcgcc aagagcgacc gcgcaagcca cacgcgccta ggccctgcgc cacggccggc 1380

cacgcgcgag aggcggggcg gagcaccgca ggagaccgac cacggacccg accggcccag 1440

ggcagcagag ccaccgagct 1460

<210> 9
<211> 738
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (287)..(287)
<223> a, c, g or t

<400> 9
agattttgaa agctcatata gaggcggcat tgggtcctcg tagtattgca tgctccgagc 60

ggacgccaag ctgtagatag gtggtggtcg cggacgaggt acttgcgtat cttgtattat 120

atgttctatc cctcaattca tacagttcaa tattccatgt gggtattgga atatgcgttt 180

catatggata tagaaacatg tctatatact atacctagtt atctcttgtc tcggatgaag 240

cttccatcta tactggctga gacagttgca gcagcagacg tcatagntat ggcgaggcca 300

caatctgacc ctcattacgt tgaacgtgca gcttatataa taaagatact gactcggtcc 360

gtgtcgcgac aaaagactca cgccgtggta aatcccagca cttgggaggc cgaggcgggt 420

tggatcacaa tggtccggag tcaaagacca gcctggccaa tatggtgaaa ccccgtctct 480

cctaaaaata caaaaattag ctgggcatag tggtgtatgc ctgtagtccc agctacttgg 540

gaggctgagg cagaagaatc gcttgaacct aggaggcaga ggttgcagtg agccgagatc 600

gtgctactgc actccagcct gggcaaaaga gcaagactcc atctcaaaaa aaaaaaaaaa 660

aaaaaaaaaa aaggcggggg gaaacccggg gccaaagcgg tcccgggggg accctggttc 720

cccgcccaaa tccccatg 738

<210> 10
<211> 909
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (623)..(623)
<223> a, c, g or t

<400> 10
ctatagtgag aaggcacttg gaaaggcaaa gtggtcttct tcttcttcct ccttctcctc 60

ctccttctcc tccttctcct ccttctcctc ctccttctcc tccttcttct cctcctcctt 120

cttctcctcc tccttctcct cctcctcctc ttcttcctct tcatcctgtt cctcttcttc 180

ttcttcttct tcttcttctt cttcttcttc ttcttcttcc tcttcttctt ccttcttctt 240

tctttttcc tttctttttt ttttgagatg gagtcttgct ctgttgccca ggctggagtg 300

cagtagcacg atctcggctc actgcaacct ctgcctccta ggttcaagcg attcttctgc 360

ctcagcctcc caagtagctg ggactacagg catacaccac tatgcccagc taatttttgt 420

atttttagga gagacggggt ttcaccatat tggccaggct ggtctttgac tccggaccat 480

tgtgatccaa cccgcctcgg cctcccaagt gctgggattt accacggcgt gagtcttttg 540

tcgcgacacg gaccgagtca gtatctttat tatataagct gcacgttcaa cgtaatgagg 600

gtcagattgt ggcctcgcca tanctatgac gtctgctgct gcaactgtct cagccagtat 660

agatggaagc ttcatccgag acaagagata actaggtata gtatatagac atgtttctat 720

atccatatga aacgcatatt ccaataccca catggaatat tgaactgtat gaattgaggg 780

atagaacata taatacaaga tacgcaagta cctcgtccgc gaccaccacc tatctacagc 840

ttggcgtccg ctcggagcat gcaatactac gaggacccaa tgccgcctct atatgagctt 900

tcaaaatct 909

<210> 11
<211> 375
<212> DNA
<213> Homo sapien

<400> 11
atctgctgct gctctgtgtc tgttctgtgc ttgcagtgct gagctggtat cacgctcaca 60

tcacatggct ggtaatacgt gtatatccac atgaatcacg gggataacag cagggaaaga 120

acatgtgaat gccaaaaggc catgcaaaaa tgccatgtgt aacctgtaaa aaaggtgccg 180

cgatggatgg agagatatat acccattagg aatcctacga gagacaataa taatagcaga 240

gagagacgga gagagaacac agacgaaaga gagagtagag acaggagaag ggaaagaaat 300

gagagaaaaa gaagagagaa cgagacaaga gaacaaagag agggcgaaac agaggcaaaa 360

aaagacaaaa aaaaa 375

<210> 12
<211> 718
<212> DNA
<213> Homo sapien

<400> 12
cggcccgggc cggtactcca tcgtcgacat ctgcctcaga tgagggatca ggcagcactc 60

taggaccaaa gaccaatctt gatccaaccc actctatact aagaattacc tcagaaccgc 120

gtgtgaatta tagactcatc cgagtagaag cgtacatttt aataggcgtg atcttggaca 180

atagactaca tccattttga ggagacatca ctatggccat gtactaaaga gactatgcat 240

gactgatgac ggaagatgtc cacggagact gtaatatacg gcctttgact atcgactaca 300

tagtaagtaa tcctgttgtc aatttgctga tgaccatgtt ggtccgagtc gcagatgcgt 360

caccgcctgt cataccagca cctaacaggt cgaggcaggc ggatcacttg aggtcaggag 420

ttcaagacca gcctggccaa tatggtgaaa cccagtctct actaaaaata caaaaattag 480

ctaggcatga tggcgcatgc ctataatccc agctactcga gtgcctgagg caggagaatt 540

gcatgaaccc gggaggagga ggaggaggtt gcagtgagcc gagatagcgg cactgcaact 600

ccagctgggt gacaaagtga gactccatct cgaaaagaca aaaccgaaag cacacacgct 660

gggggaacac actggccata atgtgtcccc gggaaaaggt atccggccaa aatcccag 718

<210> 13
<211> 686
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (285)..(285)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (296)..(296)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (312)..(312)
<223> a, c, g or t

<400> 13
agaatatacc aataggcgac ctggttcctc tagtatgcat agctcgagtc ggtcgccagt 60

gtagatagga gtaggtcgcg gacgagagta cttgcgtatc ttgtgttatc tgttctatcc 120

ctcaatatca ttacgtcaat ttccatgtgg gtatggatat gcgtatcata gcatatagaa 180

agatgtctca tatacgtata gcctaagtta tccttcgttt ctcgactgaa gctattccat 240

ttctatctac tggctgagaa cagttgagct agagaccgtc atagntagtg gcgagngcca 300

acaatctgac cnttcattac ttgacgtgca gctatataat aaatgatagc tagacctcgg 360

tccgttgcgg acgaagagag ctcagagcgt gtaaataccc agcacgtttg gcgcaggccg 420

aaggccggcg tggatcacaa tggtcacgga gcttaaagac cagcctggcc aatatggtga 480

aaccccgtct ctcctaaaaa tacaaaaatt agctgggcat agtggtgtat gcctgtagtc 540

ccagctactt gggaggctga ggcagaagaa tcgcttgaac ctaggaggca gaggttgcag 600

tgagccgaga tcgtgctact gcactccagc ctgggcaaaa gagcaagact ccatctcaaa 660

aaaaaaaaaa aaaaaaaaaa aaaaaa 686

<210> 14
<211> 720
<212> DNA
<213> Homo sapien

<400> 14
tagatcatat ggggcacatg ggtcatctag atgcatgctc gagcggcgca gtgtgatgga 60

tcccatctct actaaaaata taaaaatcag ccgggcatgg tggcatgtgc ctgtaatccc 120

agctactcag gagtctgagg aggagaatca cttgaacctg gaggcagagg ttgcagtgag 180

tcgaggttgc gctactagca ctccagcctg gacaacagag ggagactcta gtctcaaaaa 240

aacaaacaaa acctaacagc tggttcaagg caccagctgg acgggtcaag tggtgggcct 300

tttctgggtc tttggaacac tatctataga aaggttgaca aatggcttgc aaagcacagt 360

gaagaacagt gaacttataa acggggatag aattaacgtg cccagctata tagcacactt 420

tattcttatg tgcacaccaa caacaaggct atgaaaattg gtatgacgat tattaatatt 480

aatggccaaa atagtgggaa cgatattggg agactcaaga aacaggggat taatccaagt 540

ggggacccat acagtgaaca agagacaaaa ggcgcaaaga ataaaaccca aaaactcggc 600

gagggacgct acagcggaga aaaaagagca agaaaaaata aagaagaaga acaacagaag 660

caggcgggcg agccaagcac ggggaacgcg gcgggaggca cacgcggggc acaagagggg 720

<210> 15
<211> 1791
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (459)..(459)
<223> a, c, g or t

<400> 15
gcggcctgcc ctgggcaggt tacacctgcg cctgcgtgaa ggggggctcc gagtggtggc 60

gctgccgtca aatgcctgcc cgtttccttt acttattact ttatgctttt tgttcttggg 120

gaaggatgtg acagactcgc ggaggtgtcc ttgcatttcc tcgccctgat tcttgtgctt 180

tctacctccg ggtacactag agagcgtatg gcttgtagtt gtttgtgtgt tttggctttg 240

ttgtttggtt cttctataat gaaaacgtgg gacaagaaga tagaaaaaaa taacttcaca 300

tcactcaata tatctcatct gaattactac gaccttcgcc accacttcta cagggttaca 360

tgttgcggat ctcagtgtgc actcccctcg taacgcgtga atattgcgtt cccttggatt 420

gatggtgatg ctgtgtttgg tgagtctaca tcgaaacgnt cacaacaatt ctctcagtgt 480

gtgtcggaga taaagtctct gtgtggattg tccgacaccc tgtgtttatc ctcggtatgt 540

tgtctcaggg gataaaaaga ctccctctgt ttttcaccag ccccggtata taaaacattg 600

gacaaaaaac aaacgaaatg acttatacaa ggcggcttga gttggccgtt gtcccatacc 660

acaacaagat gttgtgtacg cataaaaaaa ctagttgtgt ctactcaccc ttcttgtgct 720

atagtatcac catataacgt tcagttagaa tatactcggc caaacacaac ttagaggaat 780

ataacctcgt cggcactatc aaaagcaaaa tttagcgggg gcaccaacaa cagggctttc 840

tcccacgccc cttatatgca aaacatttga ttcctttcct tttaaataac ggatgtggat 900

ttgtgtagca cttctatcta ggcatattga agttagcgaa gtgcgcgtaa gtgggtgcgt 960

gaaacaaaac aatatatata tagcaacgtg aggtccaccc ttaactatag acaacactat 1020

ttctaatatt cacaccagca ggtaacatta aacaccgatt tcatttatcc cgtaggaaag 1080

tactaccaac attacaaacc cccaacgacg acccttgagt gaccaacggt ctaaatagga 1140

atgtgaggcc cccaaaagga tcaggttgcc catggtaaga gaaaaacaac aaccgaaggc 1200

accttcccac attcgtggtg catgtgaaaa tcttatggtg acttaacacg gctaaacatg 1260

tggaccacag ccacaaagac ggaaaatatc aaatattgtg ttctacaata tagccccttc 1320

cacatgtggg tgtgaaacac atcaagccat taaaaccccc ctgtgaaaga acacttcata 1380

tacagcttaa gttgtgagtg tgcaagaaaa ccaacttata acttctgaca atatgatgtc 1440

gcacaaaaaa acattcttat aaggaccaaa agtgataata cacttcaccc aaaattataa 1500

aacgcttacc cggacaaatc ttaccccact tctatcaaaa ctattaaatg cggcaatgtg 1560

acaaacaccc ataataaacc caccacacaa aacacatatc tgacattagc ttcacctagt 1620

aacttaccac tgatcgaaag gcatatgata ctcgcatctc aaatatatct actttataaa 1680

caagatacga atatataata acatacaaca caacaaaaca aaaaaccaac aactagttga 1740

atcaacttac caaacactca tcatcggtaa aaattatcac tcaccaaacc a 1791

<210> 16
<211> 613
<212> DNA
<213> Homo sapien

<400> 16
gcggcgccgg gcaggtgcca gtgcagcgcg ctccgtgctc gaggggggcag ggggagctgg 60

aggaaaccgc agatgagttt acctctcttc gaaagataga gataaataca agctacttaa 120

aaaatatcgt caaaaggttc gctcagcatc atagctccag cgataccagt tgtgttagcc 180

gctcagatgt acacatagcg ttcaagcatg tttcacacga tgcaaaatca tgcaatgcac 240

tgtgcaggaa gccagtagcc atgcagggac ggcacagagc atcaccagag gttgcctgag 300

agagaccatg cacgggcagg ctcgcagatc gcgcaggcta ggcggtaagt catggctaca 360

tagctgactc tccgagagct ggaaagtaag taaatccgag tgcaacaaat gccgagcgac 420

aagagtaccg agcacaaata gcatgaaccg aaaagagaat accacgtacc aacccatatg 480

acaacatcac acctacataa aataatgaca ggggctgaaa caaagcgtca ggatcccaga 540

acaccataat aagcaaggag aacccagacc ccaaacaaca cacaccaaaa caaacaaaac 600

accaaaaaaa aaa 613

<210> 17
<211> 167
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (92)..(92)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (99)..(99)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (160)..(160)
<223> a, c, g or t

<400> 17
actcctccaa gaggcgacaa gttcaaagct gagtaaaggg gggaaatgaa ggaaacttct 60

tgcacaagga gcttgcccaa gctttttgtg gnggggggang aaaagtggat tgaagggagg 120

ggggcttgta aggaaagcct tgatggggcc agcccttggn attgaag 167

<210> 18
<211> 484
<212> DNA
<213> Homo sapien

<400> 18
gacaatgaaa tcatatgggc gctgggttat aatgcatgct cgagcggccg cagtgtgatg 60

gatagcggcg ccgggcaggt acactgcatt tgaatgtggc tcattgcatg gtggctcatg 120

cctatgatcc cagcactttg ggaggccaag gccagcagat catttgagcc caggagttcg 180

agaccagcct gggcaacatg gtgaaacctt gtttctatta aaaatacaaa aataaaataa 240

taattagcca actatggtgg tgcacacctg tagtcccagc tactgggggc gctaaggtgg 300

gaggatcgct tgagcctggg aggtcaaggc tgcagtgagc tatgatcaca ccactgcact 360

cctgcttagt gacagagaga gaccctgtct caaaaaaaaa aaaaaaaaaa gacttcacat 420

tcattctttc gaatttttcc ataaccccct ttagctggta taatggacag ctcttgggac 480

aaca 484

<210> 19
<211> 906
<212> DNA
<213> Homo sapien

<400> 19
tatcactatg ggcgactggg ttatctagat gcatgctcga gcggcgccag ttgtgatgga 60

ttggtcgggg ccgatgctct cgtggagccc ccttagaacc atgcgacgga taccgatgct 120

tctgtggtag actctctact gtgcgccagg ttcccatgcc gcctcctaat gccgctcgta 180

gaccttcctg actagagctg gcccattcta cttctgccat gatcaaccgc ctcccataga 240

gagaatttct atgccacact cgccatagat ttcgaacgac gtatttgtgg tgtacatcct 300

gtctgtatta atataaggac ggcgaggaca tgtcgatgga ctctgagccc acaaccgaag 360

ccatcagcaa tccccccttgt gaccttgcac gggttatcat gcccgaagca tggagacctt 420

gagagtgaga cgtcgcatag ggagagttcg accatatata cgcacacgct agaactagtt 480

gctctagcac cacagctttg agacgttgga tgcaactcgt gttccatccg actagaaatc 540

acatgtgttg atggccggtc cctggctata agtgcatgac attatctttg tcgaaagtcc 600

ttttcctggc tcatcctggc tcaaaaatca cccccactga gcaccttgca acccccccact 660

cctgcctgcc agagaacaaa ccctcttttg actgtaattt tcctttacct acccaaattc 720

cttataaaac ggcccaccct tatctccctt cgctgactct cttttcggac tcagcccgct 780

gcacccaggt gaaataaaca ggcacgtttg ctcacaaaaa aaagaaaaaa aaaaaaaaag 840

tgggggtaac cagggacaaa aggtcccggg ggaattgtga tccggccaaa ttcccaaatg 900

gacaac 906

<210> 20
<211> 744
<212> DNA
<213> Homo sapien

<400> 20
aggacaattg aagtcctata gggcgcatgg gtcctctaac tgctgctcga cgcggcgaca 60

gtgtgatgga tggctggtgt gatcatggct catgcaacct tgaattcctg ggcacaagtg 120

atcctcctgc cttagcctcc cagagtagag ttgggactac aggtatgcgc caacacacct 180

ggctaatttt attaactttt acttttagta gatgatggtg gggtgcaggt ctcactatgt 240

tgcccaggct gttctcgaac tcctggacca caagccatcc tcccacactt agtctcccaa 300

tgcgccggac ttacaggtgg ctcagtgtgt gagcacaacg tgcctggact ttactccact 360

atcttgaaat cagctgggac ggaggctttc tatctgggtg gcgactgagg agtgccaccc 420

tgaagtcacc ccaggtcatc gtgtggactg ggacatagcc taattacacc caccctgtgg 480

tagtctgtgg acagaagctt tggctgataa tcagtgggtc actacgctga taaccctgct 540

ggtgacacat tgtgattcgg ctacacacat gtccacacac aatagagaag caagagagaa 600

gacaagagag gaagagagag aagaaaaaag agaagcaaga agaaaaagaa aaaggacaga 660

cacgaggagc agcaagaaag aaggagagag cgagcagaga aaggagagag ggacagaaga 720

caggcaggga aagaagggaa gaag 744

<210> 21
<211> 851
<212> DNA
<213> Homo sapien

<400> 21
ctctctctct ctctctctct ctctctctct ctctctctct ctccctccct ttgctttccc 60

tctcctcctt tacagatgcg aaagctatag ggactatagg gcctctcgga gatagaagtg 120

ggaggagagg ataaaaaaga agagtccaga accgcgcagc agcggcagca ggacagcagc 180

aagcaaagtc gacgtgatgc gcgggcagcg agcgcgcacc cctgatgctg cagaagcaga 240

acaccagaag cggcgggggt gagcatcaac gagagcagcc catggacaaa acagccagcc 300

ttggaggaag ctgcacgacc ccgagagcac ctcctacatt caccgtgcga ggagagctga 360

cagcacagaa agtacatcac aaaagccaat cttcatctca tcgaccccgc agagcaatac 420

caggtggggg aacgaaacgc aagaagagag acgcacaagc agcagacata tcacacgccc 480

gaactgaaca tcatcaagat actcgccagg acgatgctga agcaccacac aaaacaccaa 540

atacaaagca cccgagaaca ccctgtcggc acacagcacc ccccctgcat ccgccggaac 600

agatgaacag agggcagagc aacacacgca gaaatgagaa caacctccac agcgaacaca 660

acgcagccta acacaatgat gacgatggac ctgatcaaga agcccacccc acaacgatgc 720

acaaaccagc caacagacct gacatgaacc cgcccaactg cgatcacgac gcaccaccac 780

acacagaaac gccgcaccac acaccctcgg aacacccctc cgacccccac acactcctcc 840

gaccgccgcc c 851

<210> 22
<211> 1129
<212> DNA
<213> Homo sapien

<400> 22
atggcagccg caagagaaaa tgaggaagat gcaaaagcag aatcccctga taaaaccacc 60

agatctcatg agacttattc actaccacga gaacagtatg ggggaaaccg cccccatgat 120

tcaaattatc tcccaccggg cgttggaccc gaagtccacg cgccggaact cgaaccagga 180

gctagaggac cccagcgcag aatccgcgga ggcgcagggc ttcatccacc gcgctctcag 240

atcgcgtccc agcccagatc tagccacgaa gacaccggct acctctgcct gggcttcgtg 300

ttccatgagc tccaggaagg cgacggtcag tgcctggctg gggacccgga ccgatgcgaa 360

ggctataggg actatagggc ctctcggaga tagaagtggg aggagaggat aaaaaagaag 420

agtccagaac cgcgcagcag cggcagcagg acagcagcaa gcaaagtcga cgtgatgcgc 480

gggcagcgag cgcgcacccc tgatgctgca gaagcagaac accagaagcg gcgggggtga 540

gcatcaacga gagcagccca tggacaaaac agccagcctt ggaggaagct gcacgacccc 600

gagagcacct cctacattca ccgtgcgagg agagctgaca gcacagaaag tacatcacaa 660

aagccaatct tcatctcatc gaccccgcag agcaatacca ggtgggggaa cgaaacgcaa 720

gaagagagac gcacaagcag cagacatatc acacgcccga actgaacatc atcaagatac 780

tcgccaggac gatgctgaag caccacacaa aacaccaaat acaaagcacc cgagaacacc 840

ctgtcggcac acagcacccc ccctgcatcc gccggaacag atgaacagag ggcagagcaa 900

cacacgcaga aatgagaaca acctccacag cgaacacaac gcagcctaac acaatgatga 960

cgatggacct gatcaagaag cccaccccac aacgatgcac aaaccagcca acagacctga 1020

catgaacccg cccaactgcg atcacgacgc accaccacac acagaaacgc cgcaccacac 1080

accctcggaa cacccctccg acccccacac actcctccga ccgccgccc 1129

<210> 23
<211> 900

<212> DNA
<213> Homo sapien

<400> 23
aagacgacaa aggaaagttg aatttataag ggcccattgg tttatcatag atgcatgctc 60

gagctggcgg cagtgtgatg gatccccggg caggtactgc acatagacag aaagaatggc 120

cgagctgaga ctccagtgta agacctgcac cattcttaaa tgatgcccag tgttgtagtg 180

aatgacaacc acagtgtgct gccaaatacc tgccaggatc ctacagaagg tgtgagcctg 240

atggatttgt catagtggaa tgaagctgcg gaagtccttg agtgcccata ccatatgcca 300

cattgatgca aactgcttgg ctgtaatcac tgtgtcatag ctgtattacc tgtgtgcaca 360

ttctgtatcc tggatcacaa ttcacatcac cacacaatac ttcgatgcca cctcttccac 420

catacacact cctcactaat ccaaccaacc acacaactct aaaaaccaca cacaactaca 480

tctcttacta cccacaccac ctcctcccta acatactaca ccacgactaa tcaacctcta 540

gtatcaaacc acttaaatca ctaacgacta catatcatct ctccttcgac aaccatccca 600

tatccactca caccagcact aataaataac aaacaaactc aacaagccac tagcacaaca 660

ccactcctca cactccaaac aaccaccaca gaaaatcaac caatacactt accacaccaa 720

cataaatacc caaaattacc actatcaacc tcaaaactta ctatcacata caaaatcaaa 780

tccacactac accaccatca accacagaac tactaagtcc acaactccta tgtacccacg 840

aagacactct tacactacac aacccatccc aacacatact acttaacact atccctaaca 900

<210> 24
<211> 976
<212> DNA
<213> Homo sapien

<400> 24
agatgctgct cgaccgcgcc gataatgtga tggattttag taaacagcaa aatattttga 60

aagctggatg cagatgctca gatgctagag ggtgaaatgg acagacttgg ctaggaagag 120

atatgtgaat gttagcagag ggacctttct ggggattaag gaatcaggaa cacaaatttc 180

tcttctttcc ttcccaccag gctccatgcc ccccttactg gaggaccaag accttgttgc 240

cttcaattta cgggatccca gtgggatcct gatattttcc atagtttctt aacaacattt 300

caagttaaat attaaaatta ttcatagggt gtggagtgag ccaagtgcaa cacattgctg 360

tcaggggtgt tggctactcc gccagctgtt gaaaaaagga gaaagaaaga gagcaaactg 420

agatccacac accccacaca gtatgaccaa ggcgccttct gacttcagga aagccaggca 480

gacggggatc cctggatgct cacagcttgg cagccgatat tcactggagc cagaacagtc 540

tgctctgagg cttgtctgca tccagaagtt gcaggaaagt tccacaacgt gtgaagactt 600

cttttgtcct ctctgtggga gagctgggga acataggatt ccttatagac ttatcctccc 660

cacctctcca atgagcaaag gctgctaaaa acttctgaag cctgaatccc aaaagctgga 720

ggctttctct ctcctcccag tgatcggagc ttctcagggt ggcggattgt ctcatggttc 780

tgggggccca aggcaggttc ccggcaggag tgagggtccc ggctcttggg agaggccttt 840

cagctcttgg ctccgggggtt ccacctcagc ggggctccct gcgctgcgtg ccggggtgcc 900

ggctccagac tggtcctctt cccccgcgct ttagctgggg actggccacg tccggttatt 960

tcccacccca aagaga 976

<210> 25
<211> 1660
<212> DNA
<213> Homo sapien

<400> 25
gccgtaccca cacctgcggt tgtgggctac tccaaacctg accggttaca tacttctcaa 60

taggcaggta ggtctcagtg aataattgaa aaatggtcct cactatgcac aggctataaa 120

gggactagta tttcagttta atcatgagaa aggctctctg ctttaagaag cactctacat 180

gcgtttccca agccaaactg ctggggctca actgtgggtt ctgctccttc ctggctgccc 240

aaactcaggc aagtgacttt atttcctcca tttcctcatc tataaaatgg gagtagttgg 300

gtggcattaa accgtctatg gaaagtcctc agcatggcgc ctggcacaaa gcaacggctc 360

cctcgcctct gttctcactc ctgttccaca gaattacacc cactcttctc tgctgatact 420

cttaatctca gatccccaat tctgtctttc aaatgtgttg tgtcaacatt tatttgcaaa 480

catgtctatt tgatttcaaa tgaaaacgct tttgagtgga tttaaaaaca aaacacatgc 540

gggggagaaa agagaggctg acagacatgt tagtaaacag cgaaatattt ttgaaagctg 600

gatgcagata gctcagatgc tagagggttg aaatggacag actttggcta ggaagagata 660

tgtgaatgtt agcagaggga cctttcttgg ggattaagga atcaggaaca caaatttctc 720

ttctttcctt cccaccaggg ctccatgccc cccttactgg aggaccaaga ccttgttgcc 780

ttcaatttac gggatcccag tgggatcctg atattttcca tagtttctta acaacatttc 840

aagttaaata ttaaaattat tcatagggtg tggagtgagc caagtgcaac acattgctgt 900

caggggtgtt ggctactccg ccagctgttg aaaaaaggag aaagaaagag agcaaactga 960

gatccacaca ccccacacag tatgaccaag gcgccttctg acttcaggaa agccaggcag 1020

acggggatcc ctggatgctc acagcttggc agccgatatt cactggagcc agaacagtct 1080

gctctgaggc ttgtctgcat ccagaagttg caggaaagtt ccacaacgtg tgaagacttc 1140

ttttgtcctc tctgtgggag agctggggaa ataggattcc ttatagactt atcctcccca 1200

cctctccaat gagcaaaggc tgctaaaaac ttctgaagcc tgaatcccaa agctggaggc 1260

tttctctctc ctcccagtga tcggagcttc tcagggtggg gattgtctca tggttctggg 1320

gccaaaggca gttccaggaa ggaggtgagg gtccgactct ggagagaggc atttcagctc 1380

tttggctcag gggttccatc cttcagcggg gcatccttgc agtctgctgc ctgggtgccg 1440

gtctccagac ctggctctct tccctcgcgc tcttcttcaa gcttctggga ctcagctgcc 1500

accgtggtgc tcttgctgga gtggcacccc atctcagagg gacacagagg atccagtgcc 1560

cttggatgtt ccaggaggag gaaggtctgt gccttcctcc ttgggggcca gcgttaaata 1620

accatcctct tgcagcactg ttgaaaagag ccagttccgt 1660

<210> 26
<211> 720
<212> DNA
<213> Homo sapien

<400> 26
gcgtggtcgc ggccgaggtt aatgtcactt caggaagcta ttggtgaagg tttaaacaag 60

gtgagagata ttattggaag ctggaagaaa ggtgactctt gtgacatagt agcagaaatt 120

ttagcaatgc tggaaattta ttttccatga aacagtggaa aataagtata gctcaactgg 180

atgatctcac taaagagatt tctaggcaat gtcaaaggtg ctatctggat tcttctagcc 240

cctatagcaa aagacaaaag gagaaaggca agcaagataa aaaattgttc gatataaagg 300

agccacaact ttttgggttt gaaaaatact ttttttcatt cctaacctct ccagacagtg 360

aatgatgcca aaattaagca atctgttcca gacagagcca atccagggaa ctctcagcaa 420

aatgatgaag atgaaaaggc atggctataa aaggctttgt taagaacagg aaggttaaat 480

acactgtgtt accaacaaac aatagggccc ctaaaaatct taatgtctca cggcagtttc 540

acatgggaaa ccaagataga ggtgggccat ctgaaagaga tttgtgggtg tgatttgtgt 600

ctgatggagt gaattataac tgtttaagag aaaccattaa tttaaaggat tagatcaggt 660

tgattggaaa ggatattgag ttaaatggtg cggcattgtg aatcttaatg cctaaaaata 720

<210> 27
<211> 708
<212> DNA
<213> Homo sapien

<400> 27
cgtggtcgcg gcgaggtcaa cttggaactc tggaaatgtg gcttcgctca ctggcgcctt 60

gagcttgggc gactgccggg tccgcgaaac ccgacccctg cagagctgac tccgggacta 120

ttttagtttc taacgtcaac ttgccccgat tcaagagggt ttgcgcaaaa aacgtagccc 180

gttgtcctcc tgctgcagct gttgttgcag ctgtgtggct gcgtttagta ggaataacca 240

actcaaattg ggaagttctt cagctcagta tccgctcctg taattagaac ttcttttctt 300

taagcgatga aattttggac agagagatct ggagtttagt ttgtgacgtc gaagaacaaa 360

ctccaaaatg taataccttg tccccatttg gggggcaaag ttgtgggcta attcaattcg 420

ccatggaagt gtcttctttt taaagtagtt tagtaggtat atgaatgtat ctgtcagttc 480

ttgagagacc tatggattta gcagagattt taacttagtg ccaaaaagtt tcatatttaa 540

aggcgaataa agcgaatatt tcttaaaaaa aaaaaaaaaa aagggaaaaa aacaaaaaaa 600

aaaaaaaaag ggtggggggc cccggggcca aagggttccc gggggaattg ttctcccccc 660

ccatcacacc cacaacacaa aaaaatgaaa aaggcacaac cggaccat 708

<210> 28
<211> 1099
<212> DNA
<213> Homo sapien

<400> 28
tttcattata tattgctcta tattctaggt cgcccacttt acacttcctt ctcatgcact 60

tggtcaatac cacgcccgct gaccacactg gcgacttccc tctctgtcgc ccctccgtga 120

agtcagaccc actctgcggg ccaagaaagg tgaccgggct tccttccggc ttgctaagca 180

gaggccggaa gcggtggttt ttagcggatt ctcgtagctg tgccgggtga gtggcgctcg 240

cgttcgggcg cgtgagaccc atcccgggga cccgtctccg cgggggcagc tggagggcgg 300

cggggctcct ggcgccggta gcgcacctgg gcaggtgtgc cagccaggtc cccggttctg 360

ggatccgagg ccatggcttg gagtggccca gacccgaact tcgctcctgt gcccaaactt 420

ggaactctgg aaatgtggct tcgctcactg gcgccttgag cttgggcgac tgccgggtcc 480

gcgaaacccg acccctgcag agctgactcc gggactattt tagtttctaa cgtcaacttg 540

cccgattcaa gagggtttgc gcaaaaaacg tagcccgttg tcctcctgct gcagctgttg 600

ttgcagctgt gtggctgcgt ttagtaggaa taaccaactc aaattgggaa gttcttcagc 660

tcagtatccg ctcctgtaat tagaactttt tttctttaag cgatgaaatt ttggacagag 720

agatctggag tttagtttgt gactcgaaga aaaactccaa aatgtaatac cttgtcccat 780

ttgggggggaa agtttgggct aattcaattc gccatggaag tgtcttcttt ttaaagtagt 840

ttagtaggta tatgaatgta tctgtcagtt cttgagagac ctatggattt agcagagatt 900

ttaacttaag tgcaaaaagt ttcatattta aaggcgaata aagcgaatat ttcttaaaaa 960

aaaaaaaaaa aaagggaaaa aaacaaaaaa aaaaaaaaaa gggtgggggg ccccggggcc 1020

aaagggttcc cgggggaatt gttctccccc cccatcacac ccacaacaca aaaaaatgaa 1080

aaaggcacaa ccggaccat 1099

<210> 29
<211> 598
<212> DNA
<213> Homo sapien

<400> 29
caccccacag gagtaactca tcaggactta ccagactgct gcttttgggc atcatctgct 60

gggttgatga tttggtttgg ccaagagtct tgccaagact ttaatctatg cctcttgttc 120

tacatgaatt cttgggaatt actcacgttc cataggaaga gtgcatcccc aggtgatggt 180

ttttggttat ggtatgatcc tttcacaccg agggatttca ttgtttaaaa cgtgtttctt 240

taaaagaagc cttgataacg agagtggggg aaggaggcag cagactttga agactgtggc 300

ctttggtgtt ctggagtagg gggagggaag gagaaacatg ttttccacat catcgcaagt 360

gtgtgccctt tgccccttt caggatcctt agagttgcct ccctccctcc accccgacag 420

ttttgcaata atgtgcctta tcagttgtga gtttacaggt gaagcaattt cccaaataaa 480

tggatgtaag tgttcaaaaa aaaaaaaaac aaaaaaaaag gctgggggaa accggggcca 540

aagcctctcc ccgggggggac attgtttccc gccccaattc aacccacaca aaccaccg 598

<210> 30
<211> 1495
<212> DNA
<213> Homo sapien

<400> 30
gcacgaggaa aatgctgttt gtattttgtg gtctaataag gagttcggga tagcctgttg 60

tatttgcctc atgccagccc ctgagctgcc ttgggagaag atgctgattg tccttgtcca 120

gagtactgct tttgcagagt gacaggctgc tgggacagat gtcctcctgt tgcatctttg 180

tggatgttta gtaccaatga tgacacggga actcacatca ctgacaccgc tcttcatctt 240

ctgttagtct cttgaagagc attttttttgt acttctttgc tgatgaccta cctcttcata 300

agccaagtga aacaagttga cgaactgcct aggacttcca cgtgttgctc acatacatga 360

tgatttctgt cacgctcttg tgttcagaca cactgacatt accatgtatg tcagacctcc 420

ttatgatcgc atgtcctgac agttaagctg attgcaaaca gactattaaa tatgaatgga 480

gcaaacgctg tatgtcatgg atatgttctg gagaaattct tacccatctg gatggggcag 540

ggcccttgac tcacctgaat catgaccagg caaacatttt atctgtcctt tctgcaggaa 600

tccgttctgt gtcatgctag gagaatgggt tcagtatatg gggccatcag gcagtatacc 660

ctctgaatgt ttttcattgt tgtatttgct tagagtaact aaacaattgt atcttttaat 720

ttatctttta attcaaagag gaaaccttgg cttctgataa ctttgttgtg ttgtatctta 780

atggcctata gctgtcatta cttcctgtag ctgcagtaca gaattgttac agacctggat 840

taatgcttcc aaagacagaa ggaccttggc acctaaactg accagccctg tgatcctgca 900

ccccacagga gtaactcatc aggacttacc agactgctgc ttttgggcat catctgctgg 960

gttgatgatt tggtttggcc aagagtcttg ccaagacttt aatctatgcc tcttgttcta 1020

catgaattct tgggaattac tcacgttcca taggaagagt gcatccccag gtgatggttt 1080

ttggttatgg tatgatcctt tcacaccgag gatttcattg tttaaaacgt gtttctttaa 1140

aagaagcctt gataacgaga gtgggggaag gaggcagcag actttgaaga ctgtggcctt 1200

tggtgttctg gagtaggggg agggaaggag aaacatgttt tccacatcat cgcaagtgtg 1260

tgccctttgc ccctttcag gatccttaga gttgcctccc tccctccacc ccgacagttt 1320

tgcaataatg tgccttatca gttgtgagtt tacaggtgaa gcaatttccc aaataaatgg 1380

atgtaagtgt aaaaaaaaaa aaaaaaaaaa aaaaaattct gggggaaacc ggggccaaag 1440

cctctccccg gggggacatt gtttcccgcc ccaattcaac ccacacaaac caccg 1495

<210> 31
<211> 546
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (501)..(501)
<223> a, c, g or t

<400> 31
gtttcctcag acagtcttcc ttgagcaact tcctaaacgc ccttcattat ccccctttcaa 60

gctcatgcct agagagcaag gagcaaagcc attagaaagg cttcatccca ccagcaggag 120

aagctaggac atcccaaagg gtccacttca tagagaggtg ccaaccccca cacgcacacc 180

aggcacacaa atgcatgtgt gcacacgcat accacaccct ccaattgtcc ccagaatggc 240

tcccttcagg gagtcatgtt accgggacac caaatgaggg cacaatatcc ttactcctac 300

agtttcctgc tcacattcgg attagagaaa tgggatgtct ctaaataatg tgtctaaaat 360

tctctataac ataagtgcat atgttacgtg aaaaaaacaa aaaaaaaaaa aaaaaaggtc 420

gggggggaacc cggggccaag gcggtccccg gggggaattt ggttccccgc cccaattccc 480

ccccattcgg gagaaacagg ntggcggccg agaaaacccg ggcaccaaga aagccggaca 540

cacacc 546

<210> 32
<211> 1778
<212> DNA
<213> Homo sapien

<400> 32
ggccgcctca gtttccctgg ccgtaaaacg cgggcgatgg cgctcgatgc cccgggagac 60

gccggggtcc gcgggcccct ccggcaccgg gtcctccttc aggttcgggc acccggctgc 120

ctccacgcgg ggcgagcgcg gcgatgggat cggtccgtgc tgcgctcagc cccgctttcg 180

ctatccctct gctccagggg cagcaaaatc agcagatgag cccccagcgc gcctgcggcc 240

cacacactgc ggccgtaaca gttaaacaga cggggaacat tcatgtgcgt agagctcttc 300

acggggcgat acagaaatgt gtcgttctaa agcgttaacg acctgcaatt tgcatatcct 360

ccaacacatg gattttactg attcaattat ggatgtgtgc aaagaatctg ctgctcgggc 420

gtggtgtcct cagtgctgag agagccccgc gacggcggtg cagctcctcc agggcgcgca 480

gccccgccac ctggcgccac cgcgagggag cgcaggccca gggcccgggc agacgctgga 540

ggggacgcgg caggtgaaga tggggcacct ccacttccct tctaaaggag cctgggaata 600

ggctaggaaa tgtccccact gtagaagaca gcaggaacat ctgacacgcg gataacctgt 660

gctgagctcg ttatattccc agcatttaat tctagccatt aggcatgcgt tagccacaac 720

ttcgcatagg aaggttattg caggtaagtt aagactcctt tccacttctt cctaagggtc 780

gctgacccat ggttctctag cttttgctgg taatcccacc cagattctca gcaccctagg 840

ttggaatgac ccctcctgag gtgggtgggg tagctgggag gcacctccct gaactgactt 900

ttctgccttt ttgaaggtgc agtggctgag cgctccttca gtgttttccg gtagtctttc 960

ttcagcccag tacgcagttt cctcagacag tcttccttga gcaacttcct aaacgccctt 1020

cattatcccc cttcaagctc atgcctagga gcaaggagca aagccattag aaaggcttca 1080

tccaccagca ggagaagcta ggacatccca aagggtccac ttctagagag gtgccaaccc 1140

ccacacgcac accaggcaca caaatgcatg tgtgcacacg cataccacac cctccaattg 1200

tccccagaat ggctcccttc agggagtcat gttccgggac accaaatgag ggaaaatatc 1260

cttctcctac agtttcctgc tcacatttgg attgagaaat gggatgtctc taaataatgt 1320

gtctaaattc tcttaacata agtgcatatg ttacgagaga tctgatccca gcctcccttt 1380

tttgttaaag tggtggtgct ttgccaccca tgtcaaaatg atctgggctt tccacatcaa 1440

atgcaaggga tgagcttgga ggacataccc cagtaccagc aatgatcttg agccaatggc 1500

agcgtcagct cacaacgggg tgattgaggc ttcttggtta gaagctttag aaacttgagg 1560

tcagtaaaac ccaaattgat cctttctcta gatcctctag atttctctag aggaaaaggt 1620

gacaaaaaca gacacttgtg ctcaccgtca ggggcacaat gctgtccgtc aatcctcatg 1680

cagtcctccc caccatttcg cacactgaca ccatcactac tttatgataa aaggagactc 1740

aggcagctta gataacagct gaaactcagc cagctatg 1778

<210> 33
<211> 264
<212> DNA
<213> Homo sapien

<400> 33
acccctgccg atccgcggtc ttgggtgacc tgcgcgtcga tagcgcgtgc tcgaagagga 60

gaggcaacct cagctccagg ggtttgaacc gacgccactt tcagttacga gtaaactgag 120

accagagagc agaggggggct tatttccctc tgccctatac tggccgtcga atgagcaaca 180

gtcacacaga gcaggcgacc tttttgtcaa aagtgtgtgg ggcggggcgc gcagtaggcg 240

ccctaaacgc tggactgaac agag 264

<210> 34
<211> 385
<212> DNA
<213> Homo sapien

<400> 34
tgggacctgc gccctgaatg aaggcttccc tggtaccccc tgccgatccg cggtcttggg 60

gacctgcgcg tccgatagcg cgtgctcgaa gaggagaggc aacctcagct ccaggggttt 120

tgaaccgacc cactttcagt tacgagtaaa ctgagaccag agagcagagg gggcttattt 180

ccctctgccc catactggcc ggcgaaaggg caagtcacac agagcaggcg acctttttgt 240

caaaagtgtg tggggcgggg cgcgcagtag gcgccctaaa cgctggactg aacagagaag 300

cttaattgat acttaagagt ggaagcttag ctacagttaa ggactccttc ctcctttcat 360

tcatttaata aagatttatt gattc 385

<210> 35
<211> 416
<212> DNA
<213> Homo sapien

<400> 35
cggccgcccg ggcaggtgtc tgtaacatcc atcaaggatt tccatagggg tgactggtgc 60

ccgcccaaga ctgcaccagt gcctgctcat tgaggagagt aactgctggc caggcagaaa 120

gaatatgggc tctgcaatga gacagacctg gaggggactc tcccgttgag cactagcagc 180

tggaggagtt gggagttcat ggctatcatg gttgtgttaa tcgattgtgg ggatgaaatg 240

tcattgtgta tggaaggcgg ggctcatggc tgattggcaa taaaatggcg gctgccgttg 300

tcattgtctc caaaaaaaaa aaaaaaaaaa aaaaaaggct gggggtatcc ggggccaagg 360

cggttcccgg ggggaatttg tttcccgccc caattccccc acaatttcgc aacagt 416

<210> 36
<211> 1612
<212> DNA
<213> Homo sapien

<400> 36
ctcctcccga ggaaccagtg gtgacagctg aggccatgtg agtaggatcc tgaatgaggc 60

tttatctctg gctgttcgtc ccatcgtcca ccgtggcacc agctccctca gccagccggg 120

atgggaccag cgactgagag agccagaggc agagaggtga gggtgaccat atcctggact 180

gtgagaggaa tgggactctg ggcctgtagc tgccaagcag gtggcaggtg ctccaggctg 240

tgatctgcac cctctgaccc ctgacattga cctcctaccc tgacccctgc ctgaccaagc 300

catgtctgaa caggaggctc aagccccagg gggccggggg ctgcccccgg acatgctggc 360

agagcaggtg gagctgtggt ggtcccagca gccgcggcgc tcggcgctct gcttcgtcgt 420

ggccgtgggc ctcgtggcag gctgtggcgc gggcggcgtg gcactgctgt caaccaccag 480

cagccgctca ggtgaatggc ggctagcaac gggcactgtg ctctgtttgc tggctctgct 540

ggttctggtg aaacagctga tgagctcggc tgtgcaggac atgaactgca tccgccaggc 600

ccaccatgtg gccctgctgc gcagtggtgg aggggccgac gccctcgtgg tgctgctcag 660

tggcctcgtg ctgctggtca ccggcctgac cctggccggg ctggccgccg cccctgcccc 720

tgctcggccg ctggccgcca tgctgtctgt gggcattgct ctggctgcct tgggctcgct 780

tttgctgctg ggcctgctgc tgtatcaagt gggtgtgagc ggacactgcc cctccatctg 840

tatggccact ccctccaccc acagtggcca tggcggccat ggcagcatct tcagcatctc 900

aggacagttg tctgctggcc ggcgtcacga gaccacatcc agcattgcca gcctcatctg 960

acggagccag agccgtcctt cttctcacag cggcctcagc gtccccagag ccgagccagg 1020

gtgtgagtgc atgtgaacgt tgagtacaca tgagtgcgtg tatgccccca ggctgggtca 1080

gctcttctgt ggattgcatg gcgtgtgatt aaaagcccat gtgttcccac acatccacat 1140

catgggaagg ttaatgtgtg cctccttgga actgggtgtt ggtgtccatg gaacttcctc 1200

tctgtatctc aggtcagtag gcgcagaaac gcctcatgat gaagattctt gagccccatt 1260

tccaagaccc ctcacatcca atcctgtcct gtaacatcca tcaaggattt ccataggggt 1320

gactggtgcc cacccaagac tgcaccagtg cctgctcatt gaggagagta actgctggcc 1380

aggcagaaag aatatgggct ctgcaatgag acagacctgg aggggactct cccgttgagc 1440

actagcagct ggaggagttg ggagttcatg gctatcatgg ttgtgttaat cgattgtggg 1500

gatgaaatgt cattgtgtat ggaaggcggg gctcatggct gattggcaat aaaatggcgg 1560

ctgccgttgt cattgtaaaa aaaaaaaaaa aaaaaaaaaa aaaagggcgg cc 1612

<210> 37
<211> 449
<212> DNA
<213> Homo sapien

<400> 37
ccgcccgggc aggtacagtc cagcctcggc tgggcatcag agggaaacca tgcaaagagg 60

gggaggggga gagggagcca atattttgaa atttattgag acattttaga acctagtata 120

tagcctcttg gtgaatgttc tgtgtgttct taaaaagtga atgtgtattc taccactgtt 180

cagtaaatgc caattgggtt aagtttgttg atagtcaaat ctatatcctt acccatcttt 240

ttgtcccatt ttttctgtca gttattgaac aagaggtgtt aaaatctcca attacttcta 300

tttctctaac actgccattt ttttctttgt ggattttgaa tttctctata tattttgtat 360

attttgaagg tcacatacat cttttgtcgt catgtattct gatgaattga ccctttttgtc 420

attattaaat gtgctttatc tctattcat 449

<210> 38
<211> 598
<212> DNA
<213> Homo sapien

<400> 38
aggagctgga gacagcccgg ccaacacggg gaaaccccgt ctccgccaaa agatgcgaaa 60

accagtcagg cgtggcggcg cgcgcctgcg gtcccgggca ctcggcaggc tgaggcagga 120

gaatcaggca gggaggttgc agtgagtcga gatggcggca gtacagtcca gcctcggctg 180

ggcatcagag ggaaaccatg caaagagggg gaggggggaga gggagccaat attttgaaat 240

ttattgagac attttagaac ctagtatata gcctcttggt gaatgttctg tgtgttctta 300

aaaagtgaat gtgtattcta ccactgttca gtaaatgcca attgggtcaa gtttgttgat 360

agtcaaatct atatccttac ccatcttttt gtcccatttt ttctgtcagt tattgaacaa 420

gaggtgttaa aatctccaat tacttctatt tctctaacac tgccattttt ttctttgtgg 480

attttgaatt tctctatata ttttgtatat tttgaaggtc acatacatct tttgtcgtca 540

tgtattctga tgaattgacc cttttgtcat tattaaatgt gctttatctc taaaaaaa 598

<210> 39
<211> 1016
<212> DNA
<213> Homo sapien

<400> 39
aacaaaaaga tcaacaagga atgattcact cactataggg ccactgttga ctctagatgc 60

atgctcgagc ggcgcctatg tgatggattg gtcgcggccg aggtacggag gaagctgttt 120

ctaatacttt tgttggttaa aaatacattc cgctgctgtg gaaagcctgt agctgccagg 180

agtgttacag agggcattcc tcccttgagt tgacatttgt gccaaaccgg cttttagggc 240

ttctccagct tttaagatgc gataatgata agatgatcat cgggaaaaca tccctcagat 300

gaaacgttca caggctgggc tctgaaaact ggcattcaga tgattgcggc tcccccatac 360

tgtaggaata ttgtcttatg gcttaaggtt gcctcgctcc tcatctgtaa tcagtgaaag 420

tttatccaga ggttaattac cggttttctg gtgggtatct gggaactgag gatgggcaga 480

ttaactgttg tatatgacca tagtaaagca aaagactgtt tgagaatgag taaacgttcg 540

ttttcctccc tacagaattt accagttact ctctgtccat tccactcact actgtcttca 600

tcacacacac agaggcaaac gtctgattca gatagccagg tgtgccatgc atgcactcca 660

ccatattgtg atctcttcga ggcttgtttc tgtcccctgt cctccacagt cctggatgtt 720

aatgaatagt tcatacatgt ttgttaactt acatatacct tccttttaga atttacagaa 780

aagatcaaaa gatacacaaa acaaacaggg ctgggcggta accagggcaa cgcgggcccg 840

gggggaagtg tttccgcccc aatcccgcat tacgtgagac accgggaatg gagatgatga 900

gcaagtaacg atgaacgaac gacgaggaaa gaagcagaaa caacaagaca acacacacag 960

agaacaaaag taaaggaggc agaggagaga agacagagag agaacaaaaa gagcag 1016

<210> 40
<211> 5872
<212> DNA
<213> Homo sapien

<400> 40
ggagccgggg acggcggcac cgggcgggta gggacaagac taccgcgcgt gcccccgcct 60

ggtggcagcc cctctctgcg tccctcgcgg cctggcaaaa ttacattcgg ccgagagttc 120

acgctgggga agctctctga atgcctctga gaagcgagat ccggcgccat ctcaccgacg 180

agcctcccct ttaccgcccc gtgcgtttcc tcagcacttt aggaactaaa gcctgtctgg 240

gtagctccct aacaggctct ggagctcaat ccctgggcag ggaaaagggg gtcctcgggg 300

ctccccgctc gctgtccttt ttctggacag gcagttcctt ggccacctgg tagggccgcg 360

ttgcctggca acggcggggt ccttcttggc tcggcggcgc tcggggcctg aggggagaaa 420

accgccgcgg agggcgctgg gggtggcggc ggcggtccgg gaggtggtcg cgcgactgcg 480

tggagcgcca gggcgtccga cctctgcacc tgagagaaga tgaacacggc cgaccaggcc 540

cgggtggggc ccgcggacga cgggcctgcg ccgtctgggg aggaggaggg agaggggggc 600

ggcgaggcgg gcgggaagga gccagcagcg gacgcggccc cggggcccag cgctgcattc 660

cgcctcatgg tgactcggcg ggagccggcc gtgaagctgc agtatgcggt gagcggcctg 720

gaaccgctgg cttggtccga ggaccaccgc gtgtctgtgt ccacggcccg cagcatcgct 780

gtgctggagc tcatctgcga cgtgcacaac ccgggccagg acctggttat ccaccgcacc 840

tcggtgcccg caccgctcaa cagctgtctc ctcaaagttg gctcaaaaac agaagttgct 900

gagtgcaagg agaaattcgc cgcctccaag gaccccacgg tcagtcagac tttcatgttg 960

gatagggtgt tcaaccctga ggggaaggct ttaccaccaa tgagaggatt caagtacacc 1020

agctggtctc ccatgggttg cgatgctaat ggcaggtgcc tcttggcagc actgaccatg 1080

gacaatcgcc tgaccatcca ggcaaatctc aacagactgc agtgggtcca gctggttgac 1140

ctgactgaga tctatggaga acgtctttat gagaccagtt acaggctctc taaaaatgag 1200

gccccggaag gaaatctcgg ggattttgct gagtttcaga ggagacacag catgcagacc 1260

ccagtcagaa tggagtggtc gggcatctgt accacccagc aggtcaagca taacaacgaa 1320

tgccgggacg ttggcagtgt gctcctggct gtcctctttg aaaacggtaa tatcgccgtg 1380

tggcagtttc agctgccgtt tgtaggaaaa gaatccatct cttcatgcaa cacaattgag 1440

tcaggaatca cctctcccag tgtattgttt tggtgggaat atgagcacaa taatcgaaaa 1500

atgagtggcc ttattgtggg gagtgctttt ggacccataa aaattcttcc tgtcaatctc 1560

aaagcagtca aaggctattt cactttaagg cagcctgtta tcttgtggaa agaaatggac 1620

cagttacctg tgcacagtat caaatgtgtg ccactttatc atccttacca gaagtgtagt 1680

tgcagcttag tagtggctgc aagaggctct tatgtatttt ggtgtcttct tctgatctcc 1740

aaagcagggc tgaatgttca caattcccat gtcacaggcc ttcactcact gccaattgtc 1800

tccatgactg cagacaaaca gaatggaaca gtctatactt gctccagtga cggaaaggtg 1860

aggcagctga ttcccatttt cacagatgtt gcattgaagt ttgaacacca gttgattaaa 1920

ctctcagatg tgtttggctc agtgaggact cacgggatag cagtgagccc ctgcggtgca 1980

tacctggcca tcatcaccac tgagggcatg atcaacggcc tccaccctgt taacaaaaac 2040

taccaggtcc aatttgttac tctcaaaacc tttgaagaag cagctgctca gctcctggaa 2100

tcttcagttc aaaaccttt taagcaggta gatttaatag acctagtacg ctggaagatt 2160

ttaaaagata aacatatccc tcaattttta caagaagctt tggaaaaaaa gattgaaagc 2220

agtggagtca cctatttttg gcgttttaag cttttcctcc tgaggatttt atatcagtca 2280

atgcagaaaa ccccttcaga agccttgtgg aaacccaccc atgaggactc aaaaatctta 2340

ctagtggatt cgcctgggat gggcaatgct gacgatgaac agcaggaaga aggcacttct 2400

tccaaacagg tggtgaagca aggcctgcag gagaggagca aggaaggaga tgtagaggag 2460

cccactgatg actcgctccc cacgactgga gatgctggag gccgtgagcc aatggaagag 2520

aaactcctgg aaatccaagg gaaaatcgaa gctgtggaga tgcacttgac cagggaacac 2580

atgaagcgag tcttaggaga agtgtatctg cacacctgga tcacagaaaa cactagcatc 2640

cccacccgcg gactctgtaa ctttttaatg tctgatgaag agtatgatga cagaactgca 2700

cgggtgctga ttggacatat ctcaaagaag atgaacaaac agactttccc tgagcactgt 2760

agtttgtgta aagagatctt gccattcaca gatcgcaaac aggcagtctg ttccaatggc 2820

cacatttggc tccggtgctt cttaacctac cagtcctgcc agagtttgat atatagaagg 2880

tgtttgctcc atgacagcat tgcccggcat ccagctccag aagatcccga ctggattaag 2940

aggttactgc aaagcccctg ccctttctgt gattctcctg tcttctaaat aatcagtgac 3000

gggaagatgg aagggcatga tgaactctgc catagaaaac ttcctccagc ctgaagagaa 3060

ggatgcactg gaggaagccg gaccctcacg agtggagaga agtccttggt gattgtaaag 3120

agggcccctg gagctcattt ctgaatcgca ctctccattt ccagagacta aaggatgtcc 3180

tttgaaatgg ctggactcag agagttggag tcgttttgag atgagcatta gccccagctt 3240

tgtaaccaat gaggaacact tacttatttt taagtatctt gacagaagca atttgaacac 3300

agtgtcccgt catttctaga aacagaatgg tctcttctag agagcttgga taaggacctt 3360

gctgggttga gttaggtttt aatccttgcc ttggtttgga actgccttcg ggctccagaa 3420

cttaaattgc ttggtccgtg gcatctgatg taccaacaga gattaaaagt gtaaagcaac 3480

acatgggctg atgttttgtt ctcagaaaat agctgctggt ctgcatccct ccattcttgt 3540

tttttatgca tatggaaaac attttcctaa aactctatat tcttaagttg aagccaagac 3600

taaaatttaa tgtgtcaaat gatctggtga ctattataat gaataattgt gacttatttt 3660

tcattctctc ctgggtcatc aggtttcctg acccaactcc ttaatccgta taaagatgtc 3720

aaatactgta gttcacccac gccacagccc tgcttcagac ttaactgtgg tagcctagat 3780

gagctatttg tacacagagg aaaaaaagat attttcctct tttagtaata agactttcag 3840

tatttttaat gttgacattt ccagatgttt catttagtat ccaggggtct gtctggagac 3900

ttctagagag ggacagctca gaagtgagac ccttgagctc tggtgctgta aacttgtgca 3960

attaagttga acagagcctg ggaatttctt tcctctgcac agtcccttga tatttggaat 4020

ccaggttctg cccccaaccc ctacccaccc agtggtctgt taagatgtct cagatggggc 4080

tgggcttggt ggctcacgcc tgtaatccca gcactttggg aggctgaggc gggtggatca 4140

cctgaggtca ggagtttgag accagcctgg ccaacatggt gaaaccccat ctctactaaa 4200

aacacacaca caaattagcc aggcatggtg gcacatgcct gtactcccag ctactcagga 4260

ggctgaggca ggagaatcat ttgaaccccg gaggtggagg ttgcagtgag ccgacattgt 4320

gccgcttcat tccagcctgg gtgacagagt gaaactcttg tctcaaaagt aaaaaaataa 4380

taatgtttaa aaatatttca atgtggagac aagctcaaaa tgaaattaga cacattccat 4440

tacccaggta aaagaagggg aagcctgact tgatagtagt attcaggaaa aaagagttgg 4500

cagttttatt tggccaaatt ccaatatcag ctcatggtac agcacaccgg gggaggggga 4560

cgggaggcga gaactaaggc tttttaagaa tgtgttgatg gaagtatgtg cctagatcaa 4620

aagaataatc cccccggact ccagtgtaag atcaattact gttggaatat tgtgttcctt 4680

tctagatatc acattttaag cagactttgg ccaagtagta cagtgtttgc aggagcagta 4740

acaagatggt gataactttg aaaatacttc tcaaaagaaa aataaaaaag aattagggaa 4800

gttcagtctg gagataattc agaaatacag atgataattg ttttcaaatt cttgaaggaa 4860

atgggagaga gaataggcta attctgtatt gcttcagaaa ccaaatggaa acaattaaat 4920

tccattagag aaacggttgg aaaatatgag gaggatttag ttcagtacga ggaagctgtt 4980

tctaatactt tttgttggtt aaaaatacat tccgctgctg tggaaagcct gtagctgcca 5040

ggagtgttac agagggcatt cctcccttga gttgacattt gtgccaaacc ggctttgagg 5100

gcttctccag cttttaagat gcgataatga taagatgatc atcgggaaaa catccctcag 5160

atgaaacgtt cacaggctgg ctctgaaaac tggcattcag atgattgcgg ctccccata 5220

ctgtaggaat attgtttatg gcttaaggtt gcctcgctcc tcatctgtaa tcagtgaaag 5280

tttatccaga ggttaattac cggttttctg gtgggtatct gggaactgag gatgggagat 5340

taactgttgt atatgaccat agtaaagcaa aagactgtta gagaatgagt aaacgttcgt 5400

tttcctccct acagaattta cagttactct ctgtccattc cactcactac tgtcttcatc 5460

acacacacag aggcaaacgt ctgattcaga tagccaggtg tgccatgcat gcactccacc 5520

atattgtgat ctcttcgagg cttgtttctg tcccctgtcc tccacagtcc tggattgtta 5580

atgaatagtt catacatgtt tgttaaataa atataccttc tttaaaattt acagaaaaga 5640

tcaaaagata cacaaaacaa acagggctgg gcggtaacca gggcaacgcg ggcccggggg 5700

gaagtgtttc cgccccaatc ccgcattacg tgagacaccg ggaatggaga tgatgagcaa 5760

gtaacgatga acgaacgacg aggaaagaag cagaaacaac aagacaacac acacagagaa 5820

caaaagtaaa ggaggcagag gagagaagac agagagagaa caaaaagagc ag 5872

<210> 41
<211> 757
<212> DNA
<213> Homo sapien

<400> 41
gcactgacca tataggcatg ggtcactaga gcagccgagc ggcgcagtgt gatggatgcg 60

tggtcgcggc gaggtacaga gtcctgttat ttttctcttc ggccctattt ggctgctttt 120

attaatgcat cagaacttta tgttataatc atatggattt atacgtaaat taagaaaaaa 180

tgtccaattt cattcagttc atatgttcta aacgtattgc tgatcattct taaatgagac 240

tccaggttta cattcttaca taaagtgcag ggatcccgaa gttagcccca aagatcccct 300

tgtccttttt cagacttgct caaatgttac cttatcagtg gggcctttcc tgaccacact 360

ttaaaagacc tcaacaccca cccatgggcc ttgtccctcc ttcccggctt catttttgg 420

catatactta tcaaatgtga acatatgatg catttgcttt atttatcatc gatcttcact 480

cactggcatg taagctctgt gagtgcaaag attttcatct agctatcttc cagaacagtg 540

tctggcacag agaaggagct ctatgaacta tgtgttgaat gaatggctat ctttgccttg 600

taaaccccat gctactggct ctctcttcag gtggctgacc actgcacccc aagcatgctg 660

gaaagacagg agtcccaagc cctcccttct gtctactcaa gcttgggtat catggtcata 720

gtgttcctgg tgaatgtatc gtcacatcac aaaaaaa 757

<210> 42
<211> 1895
<212> DNA
<213> Homo sapien

<400> 42
agttttgcgt tgcgccttgt tgcttgcgtc cgtcgtttgt ttgcctgtgg cttctgccgt 60

cttttttgtgg gctcgccgtt gcccgtgcct gccgctgtca cccttggcgc ccgcctgggt 120

gctggggtcc gcgattgcag tcctttgata gtgttagtga ggggggctgt cgtgcgtgtt 180

gtgtatggtc cgcatggggg agtcattagc atgttgagtt gactgtctcc cggtccgttt 240

aacgtgcgtc tggaaggtac atttttgtaa atcaagtagt tggaactaaa tccaacactg 300

ataattgcca tttcaacact gatctgaaaa gtgaattaga agctgtacaa tatcatcatt 360

agaaattctg catatggcta ataaatattc cttttaaaat taatagagtc taaagtcttc 420

caaatgatct ttacagatag agtgggacac tatagaattc tgattatatg atttagattt 480

tagggatgtt ttaacatttt caaaccacta gaaggacatt gggaacagaa agtaatagag 540

ccaacgtcac gtggtaatga tcaatagtcc agttctacga ggagaacaat tttaagctct 600

tcactgaggc caattctgct gtattctaat tccttttagg ttcttggtgg tagagtaatg 660

agctatgacc atctctggaa tactggtgag gaaaatggca gcagtaaaga aatgaggaaa 720

atattaccta attaatgata aagttaggtc cagtacagag tcctgttatt tttctctttg 780

gccctatttg gctgctttta ttaatgcatc agaactttat gtataatcat atggatttat 840

acgtaaatta agaaaaaatg tccatttcat tcagttcata tgttctaaac gtattgctga 900

tcattcttaa atgagactcc aggtttacat tcttacataa agtgcaggga tcccgaagtt 960

agccccaaag atccccttgc ctttttcaga cttgctcaaa tgttaccttа tcagtggggc 1020

ctttcctgac cacactttaa aaacctcaac acccacccat gggccttgtc ctccttcccg 1080

gcttcatttt ttggcatata cttatcaaat gtgaacatat gatgcatttg ctttatttat 1140

catcgatctt cactcactgg catgtaagct ctgtgagtgc aaagattttc atctagctat 1200

cttccagaac agtgtctggc acagagaagg agctctatga atatgtgttg aatgaatgac 1260

tatctttgcc ttgtaaaccc catgctattg gctctctctt caggtggctg accactgcac 1320

cccagggcat gctggaaaga caggagtccc aagccctccc ttctgctcta ctccaagctt 1380

ttcttcttgg gtgcattgac tcaagtcagg tagtacttct ctatgtctga gcacagacgg 1440

gctgtgttca tgtatttgta catatgtgtg aatagacaga gaaactagta gcatgggtat 1500

gtgggggaat ccatctttta gggagagatt tatctactgt ttttgtgttt agtctcacct 1560

cagaccaggt taagctggcc agggctcata gttttcaaag agcaacagaa aaaatctgtt 1620

tagcttacat tctaagcatg ttctctttat ctttcctgaa agctatccac ttttaatttc 1680

atctcatact acagagaaaa tattatttga aactgatagc tttccagaag gttactgaaa 1740

tcacttattt ttcagtgtct tcactggcac cattcatagt agctaacatt agccactttc 1800

cgtgggcctg gtgctgtgtt aaagtgcttt acatatatta tttcttttaa tccgcacaat 1860

gatcctttca agtaggtact gttattattc ccaat 1895

<210> 43
<211> 674
<212> DNA
<213> Homo sapien

<400> 43
gccttttgtg atggatgagg cggccgaggt tcgaccggcg agggaggaag aagcgcgaag 60

agccgttaga tcagtgccgg atgtggtgac ggcgtgggag actgcgggcc cgtagctggg 120

atctgcgagg tgcaagaaag cctttgaggt gataggtgta tgaaatgtca tcataacaga 180

tgtaaccaaa aacttgataa aaggttgtga aaaaactact aggatcacgc ggcatgtatt 240

gagcatatag gttgctgtag atgaatgttc ttagctgtca tgtttaaaaa tacttctgct 300

tcgttacctc aagtgtggca tgcagcattt tggaaggaaa attgaagacg tgttcaagaa 360

aacatgaaca gaagcaaatg atgaaaatga gcattttact tgacgttgat aacatcacaa 420

taaattataa agaaaaaaaa aaaaaaaaag gctgggggat aactcagggc tcaatagcgt 480

gttcccgtgg tgtgtgacaa ttgggctata ctccgcggcc tccacaaatt cccccacgac 540

caacattgcg ggagacacca aaagagaaaa caggaagaag caaaagcaca aaggccaaag 600

cagacaccaa gacaaacgaa gaagaaaaca ggggaacaaa caaaagagaa gacaaaaacg 660

aaaaaaaaga gaaa 674

<210> 44
<211> 323
<212> DNA
<213> Homo sapien

<400> 44
cgaggtaatg cttcggtgtg atgacagcgc acgttaacct cgaattcctg ggctcaggtg 60

atcctcccac ttcctgcctc ctgacttggc tgaaactaca ggcacccgcc tccaccgcca 120

gggcccagcc cacagctcct ttgacctcag tgacaggcac tcaccgtacc tgaccgccca 180

aactgaagcc tcacatttgt cccagcacgt gcccgacacc ctcatgggct accccattga 240

ccatgacaag tattccctct gctccaggga gaaaagccta ggtcccagac ctgacccatt 300

aaaacccaat cattccagct ttc 323

<210> 45
<211> 568
<212> DNA
<213> Homo sapien

<400> 45
agcgggtggg ttctgggcca gcccagcctg gaggaggtgt gagaggctga gccactgctc 60

agcttagcgg ggggaccact tagtgaccaa caccctgagg gaggccccag catcccctac 120

ttagcttggc agcagcagcg ggataaatag gggggcactg ctgcctgtga gccagcccag 180

catagccatg ggtgtgtggg ggaagcagac agagacaggg tcttgctccg ctgtccaggc 240

tggaatgctt cggtgtgatg acagcgcacg ttaacctcga attcctgggc tcaggtgatc 300

ctcccacttc cgcctcctga cttgctgaaa ctacaggcac ccgcctccac cgccaggccc 360

agcccacagc tcctttgacc tcagtgacag gcactcacct acctgacccc caaactgaag 420

cctcactttt cccagccgtg tccacaccct ctgggctacc ccattaccat gacaagtatt 480

ccctctgctc caggagaaaa gccaggtccc agacctgacc cattaaaacc caatcattcc 540

aaaaaaaaaa aaaaaactct ccagcgct 568

<210> 46
<211> 800
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (749)..(749)
<223> a, c, g or t

<400> 46
cgtggtcgcg gcgaggtgga gagcagagct gctctacccc cacctgctcc gtgttggccc 60

cagacccaac tcaggccagt ggacgtctgt cctagggctg cgtgtgagat cggtgggtgc 120

agggaacaca gcaggaagct gtgcctagaa gaaggggtc agggtcagtg tgagatgctc 180

ctaccttcag gtaaatgccc tcattctgtt agtggccacg ttcgcagcgg gctttcttgt 240

catttaccac caaaccgaac tcctagacac ataaatacac aggggatcta cccccaaaca 300

ctggcgacac taccgaacaa cggaagcagc cagggccagt agctggacag gcacaatctc 360

agcctggcag gcagcagtgc aggctgcgcg ggcctcgtgg ttggcgcagg gaggcaattg 420

cttcctgcaa tcttacttcc ctgtgtcttt ccactcggcc tgagcccgtg ccagtctggg 480

gactctgccg gccagccagc acaactgcgt ctcgccccag gggcctgact cgttgaagga 540

atgaatgaac tgactgattt tgcagggatg ggaggctggg agactgagag gtttcattta 600

aacgggaaac gctcagatct ggtggtttca caacataaaa aaaaaaaaaa aaaaacaaaa 660

aaaaaaaaaa ggctgtgggg gtaacccagg gccaaagcgc ggtttcccgg gggtgaaaat 720

ttgttttccc gccccccatc tcccccatnt ccccgaacaa acaaaacatg aggaagaaac 780

aagcaaacaa gacaccaaga 800

<210> 47
<211> 810
<212> DNA
<213> Homo sapien

<400> 47
gaaaaaaaga acagaaagaa aagagaaaag acaacacgac agaattatac atataggggc 60

ctggggtatc tagatgcatg tcgagcggcg catgtgtgat ggatgcggcg cccgggcagg 120

tactgggaaa cacggcaggt ttggtgtggg gtgggaaaat gttgaaacca cttgtagcaa 180

gccacaccta ttgggggaga ggtgataaga ttattaagct gaagtaaggc tgctagaatg 240

gcctatagaa accgcacttg gacagtggat ggcagcagaa ggccatttca ttaacagatg 300

ctgctggcag ttttgtcctg atggttggaa tccttcacca agtaatttgt atctaattac 360

aaattgtttg tatctgacac atcaatcatg attttactca gcaggcacaa cagtcaagga 420

aacacaacaa cacaccacaa caaaaacaca aaaacgcgcg ggggggacac cccagggacg 480

acgggatgga tcccggggcg gcgaactcgg tcaccccggc gccaaaattt cccaaccaaa 540

accatcggcg acaaaacggc caaggaagca agagaaacaa gaggaaacaa gaggaagaag 600

gacacggaaa gacgaaaagg agcaagaagc acgcggacaa gagacgaaga gggaggaggg 660

cgagaagagg agagagggag aggagggagg gagagacgcc aagaggggga gcgggggata 720

gagacagggg gggaggagga gagaaaaaga ggaggaaggg ggaggggggg agagcgaagg 780

ggaggaaaaa aggaagacgg agggcccgaa 810

<210> 48
<211> 818
<212> DNA
<213> Homo sapien

<400> 48
ggtcgcggcc gaggtggtgg agttgtttga aagtgacaca gcagcagtag aagcagtggt 60

gggcgaagcc caggtgaccc tcagaacgtt gcacaagaac atcagggaaa agaaccagaa 120

tcctttaagg aaaatgttct tcatgtatga gagactaaag tgatttttct aagaaagttc 180

agcccttctc tgacttacct ggacatttct agatacttcc aaaggaccct ctgggaatcc 240

atagcttcct aatctggaga tgggaggtca taagggagac gctgtggggt tccttgaagt 300

ttcttgggtt cacagaggag ccccctccac ttggtgttct cccgtgagcc agcctccacc 360

tgccaaagac actctggtcc tcgtatagtg agtaatgggg ctcagggcct ctccaacaac 420

agagaggagc tgatgctgta gggctgaccc cgtgacttcc tgagtcctca ccctgtccag 480

tgctttgaga ttcttcccac ctccccatcc tcaccagccg gatcgggcgc tgtgcagtgt 540

ggtcagcatt gggtgaagaa agtcatttcc tcgttggggc aggtattcct ctttatctct 600

cattacactg gaaatgttta tttctgctgt atcatccgtg ctcaaacgtt taagttctgt 660

caggctcacc ttctctctgg aaagaatttg cttaacttga cattccatgg tgcccgctaa 720

taaaatatat tttgaaccaa aaaaaaaaaa aaaaaaacgc tggggtaccc gggcaaaacc 780

gtcccggtga aatggtcccc cacaccaaaa aaaaaagg 818

<210> 49
<211> 1691
<212> DNA
<213> Homo sapien

<400> 49
gctgtagctg ctctgtgaaa ggtcaggcct gcccctcatg aggctccctt tatcctccta 60

aattctgggg catctacatg acgctttcta gtccaccttt gcctccgcag atcatggcta 120

ctaacctgac ctttgtctgt acttgagcac ccttcgcgat ttaactttca tgtagcgtcc 180

gacttctaat atggatttga atttcttgac tgttactgct cagaacaatc acccttttg 240

agcaggagct ggaggttatg ccgacaatga catcggagcc gtctcaacca cagggcatgg 300

ggaaagcatc ctgaaggtga acctggctag actcaccctg ttccacatag aacaaggaaa 360

gacggtagaa gaggctgcgg acctatcgtt gggttatatg aagtcaaggg ttaaaggttt 420

aggtggcctc atcgtggtta gcaaaacagg agactgggtg gcaaagtgga cctccacctc 480

catgccctgg gcagccgcca aggacggcaa gctgcacttc ggaattgatc ctgacgatac 540

tactatcacc gaccttccct aagccgctgg aagattgtat tccagatgct agcttagagg 600

tcaagtacag tctcctcatg agacatagcc taatcaatta gatctagaat tggaaaaatt 660

gtcccgtctg tcacttgttt tgttgcctta ataagcatct gaatgtttgg ttgtggggcg 720

ggttctgaag cgatgagaga aatgcccgta ttaggaggat tacttgagcc ctggaggtca 780

aagctgaggt gagccatgat tactccactg cactccagcc tgggcaacag agccaggccc 840

tgtatcaaaa aaaaaagaaa agggaaaaaa gaaagaaagc agcagcatga tcctgacatg 900

acagatgtgg gagacccaca gcctgcagac actgtgggct ggaaggtggg aagggagggg 960

ccggtggagg tggagctgtt tgaaagtgac acagcagcag tagaagcagt ggtgggcgaa 1020

gcccaggtga ccctcagaac gttgcacaag aacatcaggg aaaagaacca gaatccttta 1080

aggaaaatgt tcttcatgta tgagagacta aagtgatttt tctaagaaag ttcagccctt 1140

ctctgactta cctggacatt tctagatact tccaaaggac cctctgggaa tccatagctt 1200

cctaatctgg agatgggagg tcataaggga gacgctgtgg ggttccttga agtttcttgg 1260

gttcacagag gagccccctc acttggtgtt ctcccgtgag ccagcctcca cctgccaaag 1320

acactctggt cctcgtatag tgagtaatgg ggctcagggc ctctccaaca acagagagga 1380

gctgatgctg tagggctgac cccgtgactt cctgagtcct caccctgtcc agtgctttga 1440

gattcttccc acctccccat cctcaccagc cggatcgggc gctgtgcagt gtggtcagca 1500

tggtgaagaa agtcatttcc tcggtgggca gtattcctct ttatctctca ttacactgga 1560

aatgttattt ctgctgtatc atccgtgctc aacgttttag tctgtcaggc tcaccttctc 1620

tctggaaaga atttgcttaa cttgacattc catgtgccgc taataaaata tattttgaaa 1680

gaaaaaaaaa a 1691

<210> 50
<211> 657
<212> DNA
<213> Homo sapien

<400> 50
gggtgctata agcatggtct taatcagctc cgaccggcgc agttgtgatg gattggtcgc 60

ggcgaggtat tgttagcatt cccattttac agtggaggaa gctgaggctc agagatgtta 120

agcaagctta gctgaatggc cacaccacca gcgaagtgcc tgagccaaga tttggactcg 180

agtccatggg acccccacgc tcgtgaggct gactgctctg ctcccactgg gtcccttcat 240

gaggtcgtcc cacagcactg ctagttccag ggcgagtgcc agcacatggc cccactggga 300

gccgggggcc tgatttaggt ctactggaaa aagtgtcacc tttggggaca ctcaaggcac 360

aggctggttg gtttcgttgc tggattttat atactcatgc cctaaccctg tgttcctggt 420

ttctataagg ccccggggca aggtgcaagg aatttgcaaa tagggcctgt atgacttatt 480

tcctaggaca cgggaagctt ttcttacctc ctttctaccc tcttctccaa cctgaactcc 540

caagtttctt ctcctgaagg tctttgcact ataagcgcca aggagcccgt gtgcgtggca 600

ggggcggctg ggagggtatc tggagaacct tagtgaggcc tctggcctag ccagaga 657

<210> 51
<211> 1244
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (37)..(37)
<223> a, c, g or t

<400> 51
tgactggagt tccatgaggg agggaaattg atgtcanagt gtcattttaa agcttaagct 60

gaaagtttat tttttaaatt ctcattcatt catttagcat atattgattg agcatctaca 120

atgtgccagt tgtagaattc catctcagaa gagacttgac ttgtggatgg tggaggggca 180

gtcctgctcg gaagcagatg atgtgaaatg ttcctttcca gtctggttca cgatgtttaa 240

cagatttgtc aggtcaccac tgtgacccca agctttgctg gcagattgtt atatagtatt 300

tactgagagc cctgctatct ggtaaaggca gttaaaaagc ctgcaatctc gactcatttc 360

cagcatgaac agactggtcc ttgctgcttt acacaataat caaagctacc ttttatggcg 420

tgctcgccac tcccaagcac tgggcgaagt gctttacccg tcttcccctc cgcgatgcct 480

catgcccact ttagcagata gtactgttag cattcccatt ttacagtgga ggaagctgag 540

gctcagagag gttaagcaag cttagctgaa tggccacacc accagcgaag tgcctgagcc 600

aagatttgga ctcgagtcca tgggaccccc acgctcgtga gctgactgct ctgctccact 660

gggtcccttc atgaggtcgt cccacagcac tgctagttcc agggcgagtg ccagcacatg 720

gccccactgg gagccggggc ctgacttagg tctactggaa aaagtgtcac ctttggggac 780

actcaaggac aggctggttg gtttcgttgt ggattttata tactcatgcc ctaaccctgt 840

gttcctggtt tctataaggc cccggggaag gtgcaaggaa tttgcaaata gggcctgtat 900

gacttatttc ctaggacacg ggaagctttt cttacctcct ttctaccctc ttctccaacc 960

tgaactccca agtttcttct cctgaaggtc tttgcactat aagcgccaag gagcccgtgt 1020

gcgtggcagg ggcggctggg agggtatctg gagaacctta gtgaggcctc tggcctagcc 1080

agagaggcaa taagcttggg gacgttccgt tctgggttct gacgttgttg gttctgacgt 1140

cgttgtgctc ttttgtaaga ggaatttcat accttggaga cgctttgtac atatttgtaa 1200

tgactttatt aaaaaactga ttgtgcactt ctaaaaaaaa aaaa 1244

<210> 52
<211> 358
<212> DNA
<213> Homo sapien

<400> 52
agaatgatca tcatataggc aatggttctc tagatcatgc tcgagcggcc cagggtgatg 60

gatcgccggg caggtagaga agcctacctg ccctaatggc tcagggctat atccacctcc 120

cggataaccc tggcccttgg gactccatca tctccttgaa gtagcactga gaatccaaga 180

agaggctccg ctgctttttg cacatgttac tgagttacat ctcaggaaga tttttaagca 240

cgaggaagga aaatacaggc ctggccaagc agggtcccct tttcggtatc atctttgttc 300

ctaataagca atcaaggggg tgggtgtgtt ggctggtaaa ggaactacta agattcag 358

<210> 53
<211> 1589
<212> DNA
<213> Homo sapien

<400> 53
tggatgtata aggcaggttt tatagaacca tttagattca acattaatgg tagagtggca 60

ttttaccaaa aaaatgggtg tatttgattt gggtgctgta aggcaatttg ctaggacatg 120

ggataatcag attacacaaa atctaggcag tgaacaagtc ttcatcctgg cccaggaaat 180

agcctattaa aaaatgctgt ccaggccagg cacagtggct cacacctgta atcccaggac 240

ttcgggaggc tgaggctgga ggattgcttg agttcaagac cagcctgagc aacacagtga 300

aaacccatct caaaaaaaaa aaaaaaaaca aaccaaaaaa actggtccag ctggtgggac 360

tttcaggatt caggactgct gggggatcaa gcccaaggat gtttttcaga gctctgtgga 420

atttaagatg ctcgaaaaga gtgtccaaga ggagtaaagg tcatgacaga atttactccc 480

aggacaatct aagttctgcc acaagtaccc gtggtgtctg ttcccacaac aatggccctt 540

tcacaagctc ttgcctcaca acccctgcag aagtccttca acaaaactaa taatagacta 600

gtgaaaccta ctcctcacat gggtaagagt tgcagtgggc aggtgaccct cctccctgcc 660

cccatccttt gcttcctcaa gctccctgcc aacctctgga tacatcaatg ggaaggaaac 720

cagggaagca tagacctata gtacaacagg ggtgtagtga ccactggacc tgatgaagcg 780

aatctgcctg aaatttaata cgccttttta tttcccttct gtgttaaatc aaatgatctg 840

ttctgcactg agccaagcaa gttacttta aaactggtgg caccactcat ttgggacttg 900

gagactgctt ttatccagaa cctgttaaga gaacagggga tttaaataca aaggaatatg 960

aaggtggtgt gccttaaaga cagtctaaaa ttaggtttta gtttgttaca ttattttgaa 1020

atattaaaca tgaaaatgtt aaatcaggat gtgtgagttt taagatgttg aacactgtcc 1080

tacatcagtg aggagggagg caataaagta atttcagagt aaaacagatt gaggatgaaa 1140

ccaagacaga agtgatgatt tggcttttta tgatttttgc tgtggaaatg gcagtcgtgt 1200

gacttttcac ttgcagttta aatgaaaggg tgaagagaag cctacctgcc ctaatggctc 1260

agggctatat ccacctcccg gataaccctg gcccttggga ctccatcatc tccttgaagt 1320

agcactgaga atccaagaag aggctccgct gctttttgca catgttactg agttacatct 1380

caggaagatt tttaagcacg aggaaggaaa atacaggcct ggccaagcag ggtccccttt 1440

tcggtatcat ctttgttcct aataagcaat caaggggtg ggtgtgttgg ctggtaaagg 1500

aactactaag attcagaagc ttgtagtctt cattattttg ttttacaggt taaaataaac 1560

cacttgactg ggaaaaaaaa atggcggtc 1589

<210> 54
<211> 554
<212> DNA
<213> Homo sapien

<400> 54
tgccgaccgc gcgtagtgat ggatgagcgg cgcccgggca ggtacctggg gcttccagaa 60

tataccaaat tcagagagcc cacattcacc cctgactgtg cttggagcaa acctgaaagt 120

tcactcccaa gaggcttgtt ccagcccatt cctttattct ggaaagtaat tcttggtatt 180

gaaacggaaa actgggacaa agggagcttg aggaaaacga aaacaaacaa tgaaactgga 240

gatatgctat ttagtcttaa cccatcacag atatgctgtt tggccttaac ccatgtggaa 300

atttgtaaac tttgtcagga tttccctgtg catggtggtg aaagtcatgt agggaagaaa 360

aaattcaccg tgtaatctct ttagaagtca caaaaaaaaa aaaagctggg cgttaatcag 420

ggccataggc tgttcccggt ggtgaaatgg ttatccgcct ccacaattcc cacacaacaa 480

gggaagccaa ggtaaggcag gaggagaaag agagagcgag ggaaaagaca aaccaaacaa 540

caaccaaaaa cgcc 554

<210> 55
<211> 2581
<212> DNA
<213> Homo sapien

<400> 55
ccccaagaag cattcagtag ccttgtgctc ctaacttgtg accttgtgtg tggcgtggct 60

cccccactaa gtgtaaattt gtgttttcaa gccttccgag gacgaagtgg taagatgaaa 120

gctggcgttg ccttcgtgtt tatccgggct tctctcttcc catccttgtg atcactctgc 180

tgaccctgcc acaccctggg gccatacaca tgactcctgt gcctctgcat ctcactggcc 240

atttcaaacc agttcgtgag cctgtgttct aaaaagctca tctccattaa ctgcatctca 300

tgtcggtcac ctctgtcttt gtttcattga ctttgctgat ttaacacctc tgagaatggc 360

tttgaaagga gctacaagta atgtcttcac tggcattctt gaagtgactc ctcttggggg 420

ttaagtccca ttggcctctg ccttcctctc ctgtgatgtt gtggttgatc tactcccagc 480

caccgggagt agccgtttcc tccctgctcc tcctagtgct tgtgagcaaa ccttaccctt 540

cctccctgtc tgagccagcc tcttcctgct gccctcctca gcttgatgat gcttcaactt 600

agaagtggtg gacccttctc ggggtcagct ctacctccat gtgacccagg tgagagggag 660

acagctttca aaaggaggct ttgccttcca gatgcatccc aaaggaaata atccattagt 720

agaggcattt tgtggaggag tgaaggtgga agccagggtg cagtagggtg aggactggct 780

gggcagagag gaaactgagg cagaagatgt ggacagggac tctgcagcga ttcatggttg 840

gtcatggccg attgatacgg gcgtgtgctg agtggcagac tcataagtgc tgagctggat 900

ttctcaccca gtccttgctt ctcctgagaa cacctgcagg aagctgggaa tgaaggagag 960

aaggcaaggg taaggcctgg aagcgaggct tctctcctgt cctgcagttt attcggtggg 1020

agaaacttga ggtagaataa aaacactgaa gggattggct gtaagggagc ccagtggtca 1080

ctgccctccc agcccatagg agtagagatc agcactcctg ggagccatgg acttcttcct 1140

ctcaatgcgg gccctggggc ggcagccagc agtggtagga aagagctctt aaccacggtg 1200

ttgggattta gggaactcag ggaactagga ttaccctagt tttcagggta agtaaggaag 1260

cttgaagagc ctgttccgga agaaaagccc agaaataggc taggtgtcga cagaggtgag 1320

ggagaagcag acgttcgccc agagatggca tggtggagcc agccccctcc acgctggctg 1380

cacccctgtg gcctgggctg ttgtggcttt accagccaag ccactttctc ccatccccat 1440

gttctcagtg atttcccaag gcatttctgg tgtcctttgg gagtggaatc acgtgggttt 1500

tgaaaggaac tgcggtgaaa aatagaccct gacctgaggg caggggccga tggggaggcc 1560

actgagccct caggagtgtc tgaaagctca gaagattctg gtctgggctc tgcgggggga 1620

agccccataa gggagcgcct cctggtcttt ggtcagcttg acagagaggc ccagggaaag 1680

tacctggggc ttccagaata taccaaattc agaagcccac attcacccct gactgtgctt 1740

ggagcaaacc tgaaagttca ctcccaagag cttgttccag cccattcctt tattctggaa 1800

ataattcttg gtattgaaag gaaaactggg acaaagggag cttgaggaaa acgaaaacaa 1860

acaagaaact ggagatatgc tatttagtct taacccaaac agatatgctg tttggcctta 1920

acccatgtgg aaatttgtaa actttgtcag gatttccctg tgcatggtgg tgaaatcatg 1980

tagggaaaaa aaaattcacc gtgtaatctc tttagaagtc acaaaagaaa aagaaggtta 2040

tcttgctccc aaaaggctgt aaaaagaata agtaaagtgg ccatagaggc ctagtcttct 2100

caggacagtg tccgggttga gagtctgtct cctgaagcgc actctgggga aaatcccttc 2160

ctgccctcct gcaggtcctt agggtcccag acccagacag tcactttctc aacagagtgc 2220

cgtcaactca gcacacactc ctctcttgag cacagagccc cagagggaga agaacaaatg 2280

tgttgaaaag aatcttatta agatgtagtt aattaaaatg taatgtattg aggggaatgg 2340

aggtgtccca ggtgagggct aagtcaggca ggatttttgg ggaaggcatt gccgaaatca 2400

ccacctgagc tcaacactgg gtgcttctgg cccctccaga gttgaggtgc catccatggg 2460

aagtgcagtc ccctgcctgg cccaggttca aagcgccaag tagccacaac tcagaatgcc 2520

tgcacgttcc cctcctagcc ttatatcttc tctctggttt cctcccacga cagtttgaca 2580

t 2581

<210> 56
<211> 929
<212> DNA
<213> Homo sapien

<400> 56
gaaaaaagac ggggagaatg atactatggc ccgaatggtg cctctagatc atgctcgagc 60

ggcgcagtgt gatggattgg tcgcggcgag gtctgtggga gcctggccta cagtgtggcc 120

ttccacgtcc accggggccc tcagcctcca gtctcagaca gccctcccag ggctggccag 180

ccagaactga tgtcaccatg cccagagccc cagctcccca tactgcagaa ctgatgatgg 240

tcatgggggg cagtggagca ggggcaggag agcaggatga gcaggaatgc aataatcaag 300

atgatccaga atgagaagga agcggaagac aaggctcagt gtgagaccag ggctcagagc 360

tcagcaaact tccacgactg gctttgaatc agaatcatta tatagcttct cagccacggc 420

ccctgggtta tacagcctta aatggccctg ccaatgctgg tcacagcatt tccctagtcc 480

tggagactcg ggaactaaaa caatcaattc ccctgagcaa taaaattatg gacagctgaa 540

caacacaaag aaaacaaaaa aaaaacggct tgggggatac ctcgtgggcc aaaagcggta 600

ccccgggggt gacagtggta acccggcccc cagatccacc caaatgagag gccacaaagc 660

tggtacagct ctcccacgaa cacgcgcccg cccagagccg cgccgcgacg ccgcgacgcg 720

agcaggccga cgcgcgagag ccgctaccgc gccgccagcg ctgacgagcc aggcaggggg 780

agagcacggc gcggcaccac gacgggcgca cgcgcggcgc gcgggcggag cagcaagcgg 840

cccggaccac ggaagaggac ggcgcggcca atgcccgcga cgcgccagac ggtagcccag 900

ggggcagcag ccgcacgccg actcgagcg 929

<210> 57
<211> 984
<212> DNA
<213> Homo sapien

<400> 57
ggcggccagc gggtgggtga ggccatcttc cccatctacc cgaggccaga ccaaccccgc 60

atgaacccaa aggctcagga tcacgaggac ctgtaccgct actgtggcaa cctggctctg 120

ctccgggcta gcacggaccc cacagcccga cactgtggga gcctggccta cagtgtggcc 180

ttccacgtcc accggggccc tcagcctcca gtctcagaca gccctcccag ggctggccag 240

ccagaactga tgtcaccatg cccagagccc cagctcccca tactgcagaa ctgatgatgg 300

tcatgggggg cagtggagca ggggcaggag agcaggatga gcaggaatgc aataatcaag 360

atgatccaga atgagaagga agcggaagac aaggctcagt gtgagaccag ggtcagagct 420

cagcaaactt ccacgactgg ctttgaatca gaatcatttt gcttctcagc cacggcccct 480

gggttacaca gccttaaatg gccctgccaa tgctggtcac agcattccct agtcctggag 540

actcgggaac taaaacaatc aattcccctg agcaataaaa ttatggacag ctgaacaaca 600

caaagaaaac aaaaaaaaaa cggcttgggg gatacctcgt gggccaaaag cggtaccccg 660

ggggtgacag tggtaacccg gcccccagat ccacccaaat gagaggccac aaagctggta 720

cagctctccc acgaacacgc gcccgcccag agccgcgccg cgacgccgcg acgcgagcag 780

gccgacgcgc gagagccgct accgcgccgc cagcgctgac gagccaggca gggggagagc 840

acggcgcggc accacgacgg gcgcacgcgc ggcgcgcggg cggagcagca agcggcccgg 900

accacggaag aggacggcgc ggccaatgcc cgcgacgcgc cagacggtag cccagggggc 960

agcagccgca cgccgactcg agcg 984

<210> 58
<211> 584
<212> DNA
<213> Homo sapien

<400> 58
tgctcgagcg ccggcattat gatggattcg cgggcgaggt acacgagtgt gtgtgggtat 60

gcatgtgccc actgagagag agtatgcatg tgtgtgcact acgaacacaa gttgctgtgc 120

tggagcagga agctcgggaa acgcgagagg agagcatgca cttttagtca tccacataca 180

ttcctatgct gtgcacacac aacatccacc cagagcctgt ctcccaaatc gatggctcaa 240

ttttctactt tcttatcgta gaccagaccc cacttagacc agccggcttc aaccgttgcc 300

tgcacactta agcatcactt gacggacgct ctgtcaacaa cactctccaa tgcaccacgg 360

cacacacccc tagcaccaac tacatcagac atctctgcac gatgaacttg ggcatcaata 420

cttcatatca cactattctc atattcaata atctccttgg gctgattcca atttcctgcc 480

agccgctgag tgctcctctg cactacaacg ccctcttcct actcccctgc tcaatacacg 540

cttggccgta cctcatggtc actcgcctgt ctcctgctgt gacc 584

<210> 59
<211> 981
<212> DNA
<213> Homo sapien

<400> 59
gaaaaccaga accacacacc gggaaaacta gagaccaaaa aactagccta taacaagaac 60

ccaaaacaag accaacccac agaaaagact acaaaaacag aagctgcaca cacacacata 120

aaaaggtgtg cacacagggc acaatgaaaa aaaaaccaga aaaaacaaac ggcccctgaa 180

agggcaccct catcctata aggcctgtaa ccggtgcacc cagagcagac aagacaagga 240

gagtgtgcta caaacatcca caggtgactc tgtgaccaca aacccaaggc tggactgcaa 300

agtgctttca cagggcccca tgagggcagc tcctcgtcat ttatattttg ctgagggtct 360

ccttgaatgg ctgcttgcat aaaagtgttt agaagactgc cgttggaatc tgaatctatc 420

tgaaatgtaa ttccatttcc tggaaatgta cacgagtgtg tgtgggtatg catgtgccca 480

tgagagagag tgtgcatgtg tgtgcatacg aacacaagtt gctgtgctgg agaggaagct 540

gggaaaggag aggagagcat gcacttttag tcatccacat acatacatat gtgtgcacac 600

acacacatcc acccagagcc tgtctcccaa atcgatggct caaagtcact ttcttatcgt 660

agaccagacc ccacttagac cagcggcttc aaccttgcct gcacattaag atcacttgac 720

ggacgctctg tcaacaacac tctccaatgc accacggcac acacccctag caccaactac 780

atcagacatc tctgcacgat gaacttgggc atcaatactt catatcacac tattctcata 840

ttcaataatc tccttgggct gattccaatt tcctgccagc cgctgagtgc tcctctgcac 900

tacaacgccc tcttcctact cccctgctca atacacgctt ggccgtacct catggtcact 960

cgcctgtctc ctgctgtgac c 981

<210> 60
<211> 657
<212> DNA
<213> Homo sapien

<400> 60
tctagatgct gctcgagcgg cgcattgtga tggattggtc gcggcgaggt tgaggcctcg 60

gttcaatgag ggccccaggc aggcgacggc cacacccagt gtaaacgctg catttctaca 120

acagccacct gtgcaggccc tgcatgctct gtaacctggg gatttggtct tctgaaaagg 180

gcaccagatg aaaaactgct cttaagcctc tgttaacgtg acacagcagt agaacgtcca 240

aggtgttgat ccttggattc atgttgtctc aacttcagag acacacatcg actccttcct 300

gaccactggg catccatccc accaggagct cctaatctga gagctgttaa gaaagtcctc 360

caaaagtgct gactgcagaa gtaggtagct tctgctcaag atgacagaac aagattaact 420

tttgtattct tcagcacctt ttttattttc cattatcaca ctttgatacc ctctaaaaca 480

tttagaacac cctttctaga acgaaaaaaa aaaaaaagaa aaaaaaaaaa aaggctgtgg 540

ggggtactgt gtggccatag ggtgttcccg tggggtgaat tgtgttctcg cccaaattcc 600

ccccatttgc acaaaaagtg agcgggaaag cacggatccc tatatgtgtg gagaaac 657

<210> 61
<211> 140
<212> DNA
<213> Homo sapien

<400> 61
ccgcccgggc aggtacttct ttttatgatt ctttccacac aaaacaatca ctttgtcgca 60

ttagtatcat acccccctatg acctggacaa atcggaaata cagtttcaat ctctttctcc 120

ttctctttaa tttataaaaa 140

<210> 62
<211> 247
<212> DNA
<213> Homo sapien

<400> 62
aattgtttaa tacagaaaga gccctaggat gagtgtcctt tcccagcact gctgttagct 60

gatgtgtgac tctgggcaga tcacgtaact tcatcaactt ctgttttgta cttcttttta 120

tgattctttc cacacaaaac aatcactttg tcgcattagt atcatacccc ctatgacctg 180

gacaaatcgg aaatacagtt tcaatctctt tctccttctc tttaatttat aaaaagcatt 240

gatttta 247

<210> 63
<211> 665
<212> DNA
<213> Homo sapien

<400> 63
tcctagtatg catgctcgag ccgcgcgtat gtgatggatg tcgcggcgag gtaccgaaag 60

tgagcggggc aggcacgcta gtcacatggg taatgtggca gggtgtcgtg tcactgtgct 120

ttggctccag ggccagagca gtctgactta gtgttgagct ccaagcatgg aacacttgga 180

gtttggttca tttttgacca gcaagcctct aaatgtggtg ccttgattac ccaccgcaag 240

ggagagtggc agttgccttt ttatgacatg ttaattccag ccaggtgagt caccaggtag 300

ctctcatcct cctgccaggc tcccgctgcc tgtcggtttg gcattgtcag actagatggt 360

gactcagtgt cattggaagg tgacagtttg aggttccaaa ccagttttct cctttaagcc 420

atttcaccct caggagtgat tcctcctttg tttggcattg tcagggaatg tgatgatcca 480

ttcaaatgac ttttggagtt ccaaatagtg tttctacttt aacttccaaa aaaaaaaaaa 540

gaaaaaaaaa aaaagggcgg ggggtaccct ggggcaatag ctgtcccggt ggtggaattg 600

tttttcccgt cccaattccc cccatttttc acaacaatgg tgagcctggt caaaagagaa 660

aaact 665

<210> 64
<211> 612
<212> DNA
<213> Homo sapien

<400> 64
ggggtggcga atgatcgaca tataggggca tgggtctcta gatgctgctc gagcggcgcc 60

attgtgatgg atgcgtggtc gcgccgaggc tttgtgttaa gcgtgaggca gagggagacg 120

ttagtccaga catttccaaa gtgtgggtgg gtccgttggt tcccgagata cttttaggtg 180

gtatggggcc tgcattaagt ggcacaaaaa tcagagcaag aaagcgatgc ccttccccaa 240

ttctctcaat ccttttttatg gccgagaaga tctcagctgg atgccaacat gttccgatgc 300

ctgtggaaga catgccgacg tctcctctgc ctagggagca ggacttgggc ttagggcagg 360

tggaaaaaat tccagacttt tttagcactg tttttgtttt aatggtatat ttttattggc 420

tactttattg tttaggacaa gtggtagtgg cattcctaat ttattggggc acctttctca 480

tataatatag tattagcgca aaaaaaaaaa caaaaaaaaa aaaaggcgtg gggggaaccc 540

ggggccaaag cctgttcccg gggtgacatt ggtttcccgg cccaaaattt ccacaaaatt 600

tgggacaaat gt 612

<210> 65
<211> 365
<212> DNA
<213> Homo sapien

<400> 65
atggtgcgga tcttggccaa tggggaaatc gtgcaggatg acgacccccg agtgaggacc 60

actacccagc caccaagagg tagcattcct cgacagagct tcttcaatag gggccatggt 120

gctcccccag ggggtcctgg cccccgccag cagcaggcag gtgccaggct gggtgctgct 180

cagtcccccct tcaatgacct caaccggcag ctggtgaaca tgggctttcc gcagtggcat 240

ctcggcaacc atgctgtgga gccggtgacc tccatcctgc tcctcttcct gctcatgatg 300

cttggtgttc gtggcctcct cctggttggc cttgtctacc tggtgtccca cctgagtcag 360

cggtg 365

<210> 66
<211> 784
<212> DNA
<213> Homo sapien

<400> 66
aagtaaaaaa acaccacgag acaggtatga tatagactca tatggcgatg gtcctctaat 60

catctcgagc ggcgacagtg tgatggatcc tgcccgggca ggtactgctg gggggggttc 120

ctgccccccc cgcgcatggt ggaggtaggc tcggaccggc ccggggtagc ttgctgcagt 180

ccttcgcgcc ctcctcgccc tccccaccga catcatgctc cagattcctg cttggattaa 240

cactgggcaa ccgtggttgg aatgtactct gcgctcacga actactgata taccaaaccc 300

tggctcacct tttctctgaa cgaaattaaa aaggacttgt gactgccaaa gaacggaacc 360

ccctagtgca tgacgacgtg cctccatgca cctggccctt cagcgatata ctgattctac 420

tgctcttgag ggcctcgttt actatctgaa ccacacgctg tggcgtacct cgagtgcgtc 480

atagctggtc atccgtggtg tgaacacttg tctatccgcg tcacacattc gcacaacaag 540

gatgacgaaa gtcaaacacg gcacgaaggg agcctttaaa cggccaggga aacagcatgt 600

gcagcttgag tgaggggtca tcacataaca agtaatatct ctacccacct gaccacacaa 660

acacacacaa caaaacacac aaaacaaaca acgcgcggcg ggaaaccccc ggggcgcaac 720

acacacagac cgccggggtc gcacaaggaa tacccgcgcg cacaaaccac aacaaacagc 780

cgaa 784

<210> 67
<211> 1068
<212> DNA
<213> Homo sapien

<400> 67
aagtaaaaaa acaccacgag acaggtatga tatagactca tatggcgatg gtcctctaat 60

catctcgagc ggcgacagtg tgatggatcc tgcccgggca ggtactgctg gggggggttc 120

ctgccccccc cgcgcatggt ggaggtaggc tcggaccggc ccggggtagc ttgctgcagt 180

ccttcgcgcc ctcctcgccc tccccaccga catcatgctc cagattcctg cttggattaa 240

cactgggcaa ccgtggttgg aatgtactct gcgctcacga actactgata taccaaaccc 300

tggctcacct tttctctgaa cgaaattaaa aaggacttgt gactgccaaa gaacggaacc 360

ccctagtgca tgacgacgtg cctccatgca cctggccctt cagcgatata ctgattctac 420

tgctcttgag ggcctcgttt actatctgaa ccaaaaagct tttgtttcgt ctccagcctc 480

agcacttctc ttcttgtgct agaccctgtg tttttgcttt aaagcaagca aaatggggcc 540

ccaattgtga gaactacccg acatttccaa catactcacc tcttcccata atccctttcc 600

aactgcatgg gaggttctaa gactggaaat tatggtgcta gattagtaaa catgacttta 660

atgagtagtg tctccttaat cgttgggatt ttactacctt tttttcaaag aaacaattga 720

tgagttgtat agctggtcag atacacatca tagtgacttc accagttagg taattatcat 780

gcgaccttgt caaaccttgc tccttaatta tgttgtgcaa gtaattaaca ctgtatctca 840

gagccaggtc ggggaatact ccttattttg gacttgtaag gcgcctttgg tgctatatac 900

cccaagtcat tgtgtctctg agaagatctg tcaactgccg ctgcggggca acaacacaca 960

caaggttttc gccgcgcagc acacataagg gggtgtccaa gagagaaaga gtcccaaaca 1020

gcaaggaccg ggtgtgtaga aggacccaaa atattttaga cacgcact 1068

<210> 68
<211> 740
<212> DNA
<213> Homo sapien

<400> 68
gactgactga tatataggcc atggtttcta atcatccgag cggcgccagt tgtgaatgga 60

tcgagcggcg cccgggcagg tcgtctaaca tggcggcggc tgcggggaga gggaagcgcg 120

tttactggag ctgcattgtg agcacaaagc gaaagccaga gggggagggc agagaccagg 180

cagccgcccc tgactggcct ccttaggccc ccctctaaaa aaaaaaaaaa atcgagccac 240

agcccacgat tttatgggat tcaatattat agtcacttgt agaatcaaac tactgaggta 300

tatcttcatc tgcaagtcag acctttatgt attaattgct ttacatcgca gagacagtgt 360

aacaccttct tgtattacag gcaggggcgt gtgctatgta tgtaagagaa aaggctctgg 420

gcagagtgca ataattcaaa atgagtaaga tcagaggtgg aacggggaga aacaaattag 480

tcgtttggta aaaaccgagg taattacgtc tgtgactatc atgttaactt gaattttacc 540

ttataaagta aaatgaagcc caaaaaaaaa aaacaaaaga aaaacaaagg cggggggggc 600

accaggggcc aaacgcgggc ccccgggggg caattggttc ccggcccaca tcccacatac 660

gccgcggacg acaccccaca caacacacac agcgcacgac ccccgacaca cgacacgcac 720

ggcccacccg acaccgcaca 740

<210> 69
<211> 1028

<212> DNA
<213> Homo sapien

<400> 69
ttggggtctg tccgctcggt taccatgcac tcgagacctg tcgagcgtcc cctcttcttc 60

cgtaggagag aagtgtgttt agaatcttaa ggtagagact gcctttccgg caggcccatt 120

ttgaatgggt cttcgatttg ctaccccgcg gcccatgcga tgggctcccc tgcgtttccc 180

tctttgtttc aattgtttag gtgtcccgcc cgagcctcag gctcagctca atcgcgagat 240

gattttctgc agcgactttt tgttcctagg ggactgtgaa ggggcggggg actgccacga 300

tttagattcg ttgggggctg ggtcctgggg agactggaga ggatggctgg gactcggggc 360

acatggagag agcgtctaac atggcggcgg ctgcggggag agggaagcgc tttactggag 420

ctgcattgtg agcacaaagc gaaagcagag ggggagggca gagaccaggc agccgccccg 480

actggcctcc ttaggccccc ctctaaaaaa aaaaaaaaat cgagccacag cccacgattt 540

tatgggattc aatattatag tcacttgtag aatcaaacta ctgaggtata tcttcatctg 600

caagtcagac ctttatgtat taattgcttt acatcgcaga gacagtgtaa caccttcttg 660

tattacaggc aggggcgtgt gctatgtatg taagagaaaa ggctctgggc agagtgcaat 720

aattcaaaat gagtaagatc agaggtggaa cggggagaaa caaattagtc gtttggtaaa 780

aaccgaggta attacgtctg tgactatcat gttaacttga attttacctt ataaagtaaa 840

atgaagccca aaaaaaaaaa acaaaagaaa aacaaaggcg gggggggcac caggggccaa 900

acgcgggccc ccgggggggca attggttccc ggcccacatc ccacatacgc cgcggacgac 960

accccacaca acacacacag cgcacgaccc ccgacacacg acacgcacgg cccacccgac 1020

accgcaca 1028

<210> 70
<211> 950
<212> DNA
<213> Homo sapien

<400> 70
gggggggagg aggatgaaga actcactatg ggcgaatggg cctctagatg ctgctcgagc 60

ggcgcagtgt gaatggattc gcggccgagg tacactggcg aatattctta tttctgcaag 120

tttgcttaga ggttggcaac tgaagctgtg caggacgatt cctgttctgt aagattagtc 180

tccagttgtc agtcaagcag ttgagtgcgg tatgtctagt gcccagtttc cctctccaca 240

ggtccccata ggctcttctt gttaacttta caatccgcga tcagagatga gatctctgcc 300

aaggcagcaa ctgcaaggac catgtgggtc aatgttacca gcagacactc aaagcccatt 360

cccatttact tcaagcaccg cttttatagg attatcgttg agagacgtgg gtcatggttg 420

gtattatgag gtgagtggtc gagtgacatt cacgatttct cgatctttct gaatgcatag 480

tggctgggag tggtggctca tgcctgtgat cccggcagtt tgcggagggc cgcaggtgga 540

cagattgttt gacgcacagg cagttcgaga ccagccgggc gtaaccatgg gcgggacccc 600

caatctctac caaaaaaaaa aaaaaaatac aaaagttgtc tgggtgcggg gtcgcatgcc 660

tgtagttccc aagttcccag ctactctact tgggaggctg aggcagaaag gatcacctga 720

gcccagggaa gggccaaggc ttgcagtgag cccttgattg gtggccactt gcactttgac 780

ctttgggcaa cagaattgag aattgagacc ctgtcaaaaa aaaaaaaaaa aaaaaaaaaa 840

aaaaggtgtg ggggtataat ccatgggcaa aaagagcgtg ttccccgggg tgtgaaaatt 900

gtgtttctcc gctcaaaatt tccccaaaaa atatttggag aaaattggat 950

<210> 71
<211> 2544
<212> DNA
<213> Homo sapien

<400> 71
gggggggagg aggatgaaga actcactatg ggcgaatggg cctctagatg ctgctcgagc 60

ggcgcagtgt gaatggattc gcggccgagg tacactggcg aatattctta tttctgcaag 120

tttgcttaga ggttggcaac tgaagctgtg caggacgatt cctgttctgt aagattagtc 180

tccagttgtc agtcaagcag ttgagtgcgg tatgtctagt gcccagtttc cctctccaca 240

ggtccccata ggctcttctt gttaacttta caatccgcga tcagagatga gatctctgcc 300

aaggcagcaa ctgcaaggac catgtgggtc aatgttacca gcagacactc aaagcccatt 360

cccatttact tcaagcaccg cttttatagg attatcgttg agagacgtgg gtcatggttg 420

gtattatgag gtgagtggtc gagtgacatt cacgatttct cgatctttct gaatgcatag 480

tggctgggag tggtggctca tgcctgtgat cccggcagtt tgcggagggc cgcaggtgga 540

cagattgttt gacgcacagg cagttcgaga ccagccgggc gtaaccatgg gcgggacccc 600

caatctctac caaaaaaaaa aaaaaaatac aaaagttgtc tgggtgcggg gtcgcatgcc 660

tgtagttccc aagttcccag ctactctact tgggaggctg aggcagaagg atcacctgag 720

cccaggaggt tgagtcttgc agtgaggctg agttcacacc actgtactcg agccttgatg 780

acagaatgag actgtctcaa aaaaaaaaaa atgtccttaa gtccatgtgg acccctgact 840

aggtttgtgc cctagacagc cgtcctctga gggcaattca ggtggtgaga ctccaggttt 900

aaatggcctc cacagaaatt tcactaacct gcctttgggt ttgaccctgt ataacccctt 960

tcttctggag gtccctttgg gtggcagtag atacgggatt tggtgtctga cagctctggg 1020

gacagatccc agctccaaat ggcagagtct ctacagatta caagccaaat acttagcact 1080

atgtgctgat cttcaggaag tcagtctata tttcataaca agtcacatgg ggataatgaa 1140

ggaatggcct aaaatgctct cagtaatatt cctgagtcat ccctcagggc taggcttggt 1200

gttaggcatg gcggggaagg gagcagagct gtgtgcagag gaagatgcag ttcttgcctt 1260

gtcagggtcc ctgacctgat ggcgacccat ggtggagtct tcatagtgac agacaccact 1320

gtaaaagcag atccaggttg tgcaaccctc aaagcaggtc tcctcactca ccgggataga 1380

tagactattg gccgtacctg catccaccgc ttgccatggt ttcgttgtgg gtggaggata 1440

ctttcctgtc ccctggcttt gggtttgccc acgtggcttg ctctggcctt ggaatgaagc 1500

agaaacgaaa ggctgccagt tccgagccca cgtctgaagt cgccttaggt ggttccgcgg 1560

gccccgtgcg ctcccacctt cacccagagg gccttctctg gtgcagccgc tgcttcttca 1620

gcctccgccc aaaaggaacg gagccccctg gccgatccgc aggcctacag ggagccacag 1680

agcgcagcgg ctggaccagc gttcaagccc aagcacaggc ctgcgagaac cttgttccag 1740

ccgccgttta ggatggttga ttaggacgcg ttgcagtggc ggtagctcac caatccagtg 1800

cgtgcacccg ctcctttatt aggctataga gccagtggct cccacaggga cctgatacaa 1860

cagtgcgtta aataaggagc atattgagct ctcatgtcgt aagccagtgg agaagtccag 1920

ggctagtgtg ggggctccgg cgggggctgt ggcccccatc cgcatggagc ctccccatgg 1980

ttcacaggtc tcagtcttcg gagccttcgg ccctgcgagc ccgaacagtc cacagggcgg 2040

cgccagaccc tctttcgaac gccatcctct aaagcctcgg ctccaaccgg ttccacttct 2100

tcaggctcag gattttcact cttctcgaat gggggtggcc ctcccccaat cttctgagtc 2160

gcaacagcat ctccctccct ccaggacctc agagccagag ctgggcgaga ggccctgacc 2220

tccggggtag ggtggaagcg tccctgtgaa ggtgcagtcc tgcctcccat ccccaggcgc 2280

cgggcctctc ccaccctcag cgccctgctc acctccagct gaagatgcca gggcacctct 2340

gcttcctccc tgccctctct gcagtaccgc cgagtgtgca taaaagggtt taatataggc 2400

tttgccgggc gcggggactc ccacctgtaa tcccagtacg ttgagagacc aaggcgggag 2460

gatcacttga ggccaggagt tcaaaaccag cctgggcaac aaagtgaggc ccgtctctga 2520

aaaaaaaaaa aaaaaaaaaa gggt 2544

<210> 72
<211> 328
<212> DNA
<213> Homo sapien

<400> 72
aggacgtgat gatcatatag gggaatgggt catctagatg atgctcgagc gtgcgcagtg 60

tgatggattt atatcttaat ttttaatcat gtcagttctt gaatgggtat ctccttagcc 120

tgctgatttc ttttttcttc taaagaaagt gggtggagaa attaatttag acgtttgttt 180

gcaataaaaa gaattcattt taaaaaaaaa aaaaaaaaaa agctgtggcg gtaatcagtg 240

gctcatagcg gttttccgtg gtgtgaaact ggttatccgg ctcacaattt ccaacacaga 300

catagcagag acaagttcca cgacaaaa 328

<210> 73
<211> 482
<212> DNA
<213> Homo sapien

<400> 73
tataaactgt tttaaaagaa acccatgaaa tttttaaagg atttgcatca ggttggattg 60

agaaggatag taggagtata aatggtgcag ccactatgga aaagtctgac agtgcctcaa 120

aagactaaac ataaaggtac cgtataccca acaattccac ccctaagtat atacccaaga 180

aaatgaaaac atgtccacat aaaaaattgt acacagatgg tgtttgtagc agcattattt 240

gtaataacca aaaagtagaa acaatgcaaa tgcccatcag ctgatgagtg gaaatgtaaa 300

ctgtgatgta ttcatacaat ggaatattat ttgacaataa aaataagtgg agtgccagta 360

catgctataa caaaaaaaaa aaaaaaaaaa aaactttggg gttatctcat ggctcatacc 420

tttttccctg ttttgacatt ttttttccgc ttccaatttc cacacaaatc ttgacacaaa 480

tt 482

<210> 74
<211> 1187
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (298)..(298)
<223> a, c, g or t

<400> 74
taaatcttga tataccattt gctgctacta agccatgtag taggagcttt gggactgagc 60

ttgagtgtaa gttagaaggg ccttgaagag actgttaacc ggagcctaat gtcacttcag 120

gaagctattg gtgaaggttt aaaccaggtg agagatatta ttggaagctg gaagaaaggt 180

gactcttgtg acatagtagc agaaatttta gccatgctgg aaatttattt tccctggaaa 240

ccattggaaa ataagtatag ctcaactgga tgatctcact aaagagattt ctaggcantg 300

tcaaaggtgc tatctggatt cttctagccc ctatagcaaa gacaaaggag aaaggcaagc 360

aagatacaaa atttgtttcg atataaagga gccacaactt tttgggtttg aaaatacttt 420

tgtgtcattc ctaacctctc cagacagtga atgatgccta atattaagca atctgttcca 480

gacagagcca atccagggaa ctctcagcaa aatgatgaag atgaaaaggc atggctataa 540

aaaggctttg ttaagaacag gaaggttaaa tacactgtgt taccaacaaa caatagggcc 600

cctaaaaatc ttaatgtctc acggcagttt cacatgggaa cccaagatag aggtgggcca 660

tctgaaagag atttgtgggt gtggtttgtg tctgatggag tgaattataa actgttttaa 720

aagaaaccca tgaaattttt aaaggatttg catcaggttg gattgagaag gatagtagga 780

gtataaatgg tgcagccact atggaaaagt ctgacagtgc ctcaaaagac taaacataaa 840

ggtaccgtat acccaacaat tccaccccta agtatatacc caagaaaatg aaaacatgtc 900

cacataaaaa attgtacaca gatggtgttt gtagcagcat tatttgtaat aaccaaaaag 960

tagaaacaat gcaaatgccc atcagctgat gagtggaaat gtaaactgtg atgtattcat 1020

acaatggaat attatttgac aataaaaata agtggagtgc cagtacatgc tataacaaaa 1080

aaaaaaaaaa aaaaaaact ttggggttat ctcatggctc ataccttttt ccctgttttg 1140

acatttttt tccgcttcca atttccacac aaatcttgac acaaatt 1187

<210> 75
<211> 759
<212> DNA
<213> Homo sapien

<400> 75
catttcctgg gcacgcatgg tggcaaaggg agaagtgacc agaatcagat ttggtggcca 60

aagggcacag ccagcatctg tgccatgcct ctgatctccc cccaccatat gaagggaagg 120

ggccagctgt atccctctgg tggcttggtg gctctccttg gaatggagag gagtctgtgg 180

ctttccatct tcctgcaaag tggctggagt tggtgtccga tagctgcaaa ctccaggcag 240

catgagcgtg ctgctgaagc taggagcatg caatttccca cagcctggag cagggatttt 300

cagactggga cctaaagtcc taggcttcat caaagtctgt gtcccatccc agttccagct 360

gcactctcag gggtttgtgt gccttactgc ttttattttc cacttgttta agtctgaggc 420

tgttagcaag ctgaattata tagcagttta gggacatgcc ctggaattag gagctggatg 480

agaatcccac ctcctctcct cactcacact atgatcttgc caattacatc acttttgaaa 540

gccctgtccc ttcttctaca aaatgggttc actagtcagg gagctgaaag gagctgattc 600

taataaagca cctagaaaca cggtcttagt gttggcccac tctgcaggtc agaggggtc 660

ctaggtgctc aggaaggctt tcaaggtaag tgtggagcac cggtgtctgc agtgagcggg 720

gagcttttgt cctgtgattg tggcagccaa accggaagc 759

<210> 76
<211> 943
<212> DNA
<213> Homo sapien

<400> 76
actagttctc ctaatattct gggcttaaac tacactggga ggggcttgca tttcctgggc 60

acgcatggtg gcaaagggag aagtgaccag aatcagattt ggtggccaaa gggcacagca 120

gcatctgtgc catgcctctg atctcccccc accatatgaa gggaaggggc cagctgtatc 180

cctctggtgg cttggtggct ctccttggaa tggagaggag tctgtggctt tccatcttcc 240

tgcaaagtgg ctggagttgg tgtccgatag ctgcaaactc caggcagcat gagcgtgctg 300

ctgaagctag gagcatgcaa tttcccacag cctggagcag ggattttcag actgggacct 360

aaagtcctag gcttcatcaa agtctgtgtc ccatcccagt tccagctgca ctctcagggg 420

tttttgtgtg ccttactgct tttattttcc acttgtttaa gtctgaggct gttagcaagc 480

tgaattatat agcagtttag ggacatgccc tggaattagg agctggatgg gaatcccacc 540

tcctctcctc actcacccta tgatcttgcc aattacatca cttttgaaag ccctgtccct 600

tcttctacaa aatgggttca ctagtcaggg agctgaaagg agctgattct aataaagcac 660

ctagaaacac ggtcttagtg ttggcccact ctgcaggtca gaggggtcc taggtgctca 720

ggaaggcttt caaggtaagt gtggagcaca ggtgtctgca gtgagcgggg agcttttgtc 780

ctgtgattgt ggcagcaaac ccggaaagcc ttgccctgca ttccctccag gggcgggccg 840

ctaggatcaa ttgttccttc ccctggatcc acttttaaag ccctacccac actgtcagag 900

gggcagagcc tgggctagca gggaaggagg ccccttcaga gtg 943

<210> 77
<211> 244
<212> PRT
<213> Homo sapien

<400> 77

Met Gly Ile Phe Leu Lys Ala Cys Leu Cys Ala Asn Pro Ser Pro Lys
1 5 10 15

Gly Gly Tyr Leu Arg Trp Val Glu Pro Ser Ser His Gly Val Glu Arg

```
            20                  25                  30

Arg Pro Trp Thr His Thr Arg Glu Glu Pro Pro Lys Pro Ser Ser Ile
        35                  40                  45

Met Trp Gln Arg Ile Gln Arg Trp Ala Tyr Leu Ser Gly Ser Ile Ala
    50                  55                  60

Cys Leu Arg Gly Ala Asp Asn Cys Arg Thr Ser Ala Ser Gln Phe Ser
65                  70                  75                  80

His Gln Thr Lys Ile Cys Asp Thr Asn Thr Gln Pro Gly Ala Ser Pro
                85                  90                  95

Thr Asp Ala Arg Lys Ala Arg Arg Pro Lys Ser Pro Arg Pro Arg Pro
            100                 105                 110

Ala Pro Ala Pro Arg Gln Ala Pro Gly Gln His Pro His Ser Thr Thr
        115                 120                 125

Gly Ala Ala Ile Thr Thr Gly Pro Thr Ala Gln Arg Arg Glu Ala Thr
    130                 135                 140

Asp Ala Glu Asn Lys Arg Lys Arg Thr Arg Gln Arg Thr Arg Arg Thr
145                 150                 155                 160

Thr Gly Gln Thr Tyr Glu Gln Thr Lys Lys Arg Lys Lys Lys Thr Lys
                165                 170                 175

Arg Asp Ala Gly Asp Asp Gly Arg Ala Arg Lys Thr Lys Arg Gln Ala
            180                 185                 190

Lys Arg Asn Lys Gly Lys Ala Lys Arg Gly Arg Ser Lys Gln Glu Arg
        195                 200                 205

Lys Lys Lys Gln Arg Ala Thr Lys Gln Glu His Lys Glu Lys Asp Arg
    210                 215                 220

Lys Ala Pro Arg Gly Gln Thr Lys Glu Gly Glu Gln Asn Thr Lys Asp
225                 230                 235                 240

Glu Arg Glu Glu
```

<210> 78
<211> 104
<212> PRT
<213> Homo sapien

<400> 78

Met Gly Tyr Pro Ala Ser Lys Phe Ser Pro Thr Thr Leu Glu Arg Gln
1 5 10 15

Gln Pro Arg Lys Gln Thr Gln Arg Ala Ser Ser Gln Arg Gln Gly Asn
20 25 30

Asn Thr Lys Ala His Arg Gln Lys Glu Gly Ala Ala Glu Gly Thr Gln
35 40 45

Ala Thr Pro Glu Arg Gly Gln Thr Gln Ala His Gln Lys Arg Arg Glu
50 55 60

Arg Thr Thr Gly Arg Glu Glu Gln Lys Glu Lys Arg Gln Gln Arg Glu
65 70 75 80

Glu Gln Gly Thr Arg Gly Asp Arg Glu Arg Lys Arg Gln Pro Ala Asn
85 90 95

Ala Gln Asp Gly Gln Gln Ala Arg
100

<210> 79
<211> 54
<212> PRT
<213> Homo sapien

<400> 79

Met Arg Val Tyr Ala Cys Ser Ser Val Tyr Ser Gln His Arg Gly Ser
1 5 10 15

Phe Asp Val His Val Tyr Leu Tyr Tyr His Gly Tyr Val Gly Val Thr
20 25 30

Thr Leu Thr Met Ile Phe Ser Ser Val Leu Phe Gly Tyr Gly Phe Gly
35 40 45

Val Ile Trp Leu Leu Leu
50

<210> 80

<211> 76
<212> PRT
<213> Homo sapien

<400> 80

Met Ser Glu Thr Pro Gly Gln Val Pro Gly Asp Arg Cys Ser Pro Ser
1 5 10 15

Pro Val Lys Val Asp Ala Leu Glu Met Glu Pro Met Ser Pro Trp Glu
20 25 30

Arg Leu Asp Cys Val Lys Leu Arg Ser Arg Asp Val Gly Arg Ser Ala
35 40 45

His Ala Ala Tyr Ile Val Pro Cys Thr His Ile Cys Ala Arg Leu Ala
50 55 60

Ser Asp Gly Asp Phe His Glu Leu Ile Glu Gly Thr
65 70 75

<210> 81
<211> 125
<212> PRT
<213> Homo sapien

<400> 81

Met Arg Tyr Ala Ala Ser Asn Ser Pro Gly Ser Tyr Arg Pro Lys Lys
1 5 10 15

Val Asp Arg Ala Ala Ala Glu Glu Gln Ala Phe Asp Gly Met Pro Asn
20 25 30

Thr Glu Gly Arg Arg Pro Ala Gly Asp Pro Gly Arg Arg Ser Pro Thr
35 40 45

Ala Ala Gly Arg Gly Glu Gly Gln Ile Arg Gly Arg Glu Pro His Ala
50 55 60

Arg Pro Cys Met Arg Arg Arg Arg Pro Arg Glu Arg Arg Pro Glu Ala
65 70 75 80

Ala Arg Gln Glu Arg Pro Arg Lys Pro His Ala Pro Arg Pro Cys Ala
85 90 95

Thr Ala Gly His Ala Arg Glu Ala Gly Arg Ser Thr Ala Gly Asp Arg
100 105 110

```
Pro Arg Thr Arg Pro Ala Gln Gly Ser Arg Ala Thr Glu
        115                 120                 125


<210>  82
<211>  235
<212>  PRT
<213>  Homo sapien

<400>  82

Ala Trp Ala Leu Leu Phe Leu Thr Leu Leu Thr Gln Gly Thr Gly Ser
1               5                   10                  15


Trp Ala Gln Ser Ala Leu Thr Gln Ser Ala Ser Val Ser Gly Ser Pro
            20                  25                  30


Gly Gln Ser Ile Thr Ile Ser Cys Thr Gly Thr Ser Ser His Val Gly
        35                  40                  45


Gly Tyr Asn Tyr Val Ser Trp Tyr Gln Gln His Pro Gly Lys Ala Pro
    50                  55                  60


Lys Leu Ile Ile Tyr Glu Val Ser Asn Arg Pro Ser Gly Val Ser Asn
65                  70                  75                  80


Arg Phe Ser Gly Ser Lys Ser Gly Asn Thr Ala Ser Leu Thr Ile Ser
                85                  90                  95


Gly Leu Gln Ala Glu Asp Glu Ala Asp Tyr Tyr Cys Cys Ser Tyr Thr
            100                 105                 110


Arg Ser Thr Ser His Val Phe Gly Thr Gly Thr Lys Val Thr Val Leu
        115                 120                 125


Gly Gln Pro Lys Ala Asn Pro Thr Val Thr Leu Phe Pro Pro Ser Ser
    130                 135                 140


Glu Glu Leu Gln Ala Asn Lys Ala Thr Leu Val Cys Leu Ile Ser Asp
145                 150                 155                 160


Phe Tyr Pro Gly Ala Val Thr Val Ala Trp Lys Ala Asp Gly Ser Pro
                165                 170                 175


Val Lys Ala Gly Val Glu Thr Thr Lys Pro Ser Lys Gln Ser Asn Asn
            180                 185                 190
```

```
Lys Tyr Ala Ala Ser Ser Tyr Leu Ser Leu Thr Pro Glu Gln Trp Lys
        195                 200                 205

Ser His Arg Ser Tyr Ser Cys Gln Val Thr His Glu Gly Ser Thr Val
    210                 215                 220

Asp Glu Asp Ser Gly Pro Leu Gln Lys Cys Ser
225                 230                 235
```

<210> 83
<211> 166
<212> PRT
<213> Homo sapien

<400> 83

```
Pro Pro Pro Ser Pro Pro Ser Pro Pro Ser Pro Pro Pro Ser Pro Pro
1               5                   10                  15

Ser Ser Pro Pro Pro Ser Ser Pro Pro Pro Ser Pro Ser Ser Ser Ser
            20                  25                  30

Ser Ser Ser Ser Ser Cys Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
        35                  40                  45

Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Phe Phe Phe Leu Phe
    50                  55                  60

Ser Phe Leu Phe Phe Leu Arg Trp Ser Leu Ala Leu Leu Pro Arg Leu
65                  70                  75                  80

Glu Cys Ser Ser Thr Ile Ser Ala His Cys Asn Leu Cys Leu Leu Gly
                85                  90                  95

Ser Ser Asp Ser Ser Ala Ser Ala Ser Gln Val Ala Gly Thr Thr Gly
            100                 105                 110

Ile His His Tyr Ala Gln Leu Ile Phe Val Phe Leu Gly Glu Thr Gly
        115                 120                 125

Phe His His Ile Gly Gln Ala Gly Leu Ala Leu Arg Thr Ile Val Ile
    130                 135                 140

Gln Pro Ala Ser Ala Ser Gln Ser Ala Gly Ile Tyr His Gly Val Ser
145                 150                 155                 160
```

Leu Leu Ser Arg His Gly
165

<210> 84
<211> 63
<212> PRT
<213> Homo sapien

<400> 84

Met Glu Arg Tyr Ile Pro Ile Arg Asn Pro Thr Arg Asp Asn Asn Asn
1 5 10 15

Ser Arg Glu Arg Arg Arg Glu Asn Thr Asp Glu Arg Glu Ser Arg Asp
20 25 30

Arg Arg Arg Glu Arg Asn Glu Arg Lys Arg Arg Glu Asn Glu Thr Arg
35 40 45

Glu Gln Arg Glu Gly Glu Thr Glu Ala Lys Lys Asp Lys Lys Lys
50 55 60

<210> 85
<211> 98
<212> PRT
<213> Homo sapien

<400> 85

Met Gly Phe Trp Pro Asp Thr Phe Ser Arg Gly His Ile Met Ala Ser
1 5 10 15

Val Phe Pro Gln Arg Val Cys Phe Arg Phe Cys Leu Phe Glu Met Glu
20 25 30

Ser His Phe Val Thr Gln Leu Glu Leu Gln Cys Arg Tyr Leu Gly Ser
35 40 45

Leu Gln Pro Pro Pro Pro Pro Pro Gly Phe Met Gln Phe Ser Cys Leu
50 55 60

Arg His Ser Ser Ser Trp Asp Tyr Arg His Ala Pro Ser Cys Leu Ala
65 70 75 80

Asn Phe Cys Ile Phe Ser Arg Asp Trp Val Ser Pro Tyr Trp Pro Gly
85 90 95

Trp Ser

<210> 86
<211> 53
<212> PRT
<213> Homo sapien

<400> 86

Met Arg His Leu Ser Ile Cys Tyr Asp Thr His Ile His Thr His Met
1 5 10 15

Glu Ile Asp Val Met Ile Leu Arg Asp Arg Thr Asp Asn Thr Arg Tyr
20 25 30

Ala Ser Thr Leu Val Arg Asp Leu Leu Leu Ser Thr Leu Ala Thr Asp
35 40 45

Ser Ser Tyr Ala Tyr
50

<210> 87
<211> 73
<212> PRT
<213> Homo sapien

<400> 87

Leu Lys Asp Gln Pro Gly Gln Tyr Gly Glu Thr Pro Ser Leu Leu Lys
1 5 10 15

Ile Gln Lys Leu Ala Gly His Ser Gly Val Cys Leu Ala Ser Gln Leu
20 25 30

Leu Gly Arg Leu Arg Gln Lys Asn Arg Leu Asn Leu Gly Gly Arg Gly
35 40 45

Cys Ser Glu Pro Arg Ser Cys Tyr Cys Thr Pro Ala Trp Ala Lys Glu
50 55 60

Gln Asp Ser Ile Ser Lys Lys Lys Lys
65 70

<210> 88
<211> 90
<212> PRT
<213> Homo sapien

<400> 88

Met Lys Ile Gly Met Thr Ile Ile Asn Ile Asn Gly Gln Asn Ser Gly
1 5 10 15

Asn Asp Ile Gly Arg Leu Lys Lys Gln Gly Ile Asn Pro Ser Gly Asp
20 25 30

Pro Tyr Ser Glu Gln Glu Thr Lys Gly Ala Lys Asn Lys Thr Gln Lys
35 40 45

Leu Gly Glu Gly Arg Tyr Ser Gly Glu Lys Arg Ala Arg Lys Asn Lys
50 55 60

Glu Glu Glu Gln Gln Lys Gln Ala Gly Glu Pro Ser Thr Gly Asn Ala
65 70 75 80

Ala Gly Gly Thr Arg Gly Ala Gln Glu Gly
85 90

<210> 89
<211> 96
<212> PRT
<213> Homo sapien

<400> 89

Met Leu Phe Val Leu Gly Glu Gly Cys Asp Arg Leu Ala Glu Val Ser
1 5 10 15

Leu His Phe Leu Ala Leu Ile Leu Val Leu Ser Thr Ser Gly Tyr Thr
20 25 30

Arg Glu Arg Met Ala Cys Ser Cys Leu Cys Val Leu Ala Leu Leu Phe
35 40 45

Gly Ser Ser Ile Met Lys Thr Trp Asp Lys Lys Ile Glu Lys Asn Asn
50 55 60

Phe Thr Ser Leu Asn Ile Ser His Leu Asn Tyr Tyr Asp Leu Arg His
65 70 75 80

His Phe Tyr Arg Val Thr Cys Cys Gly Ser Gln Cys Ala Leu Pro Ser
85 90 95

<210> 90

<211> 91
<212> PRT
<213> Homo sapien

<400> 90

Met Gly Trp Tyr Val Val Phe Ser Phe Arg Phe Met Leu Phe Val Leu
1 5 10 15

Gly Thr Leu Val Ala Arg His Leu Leu His Ser Asp Leu Leu Thr Phe
20 25 30

Gln Leu Ser Glu Ser Gln Leu Cys Ser His Asp Leu Pro Pro Ser Leu
35 40 45

Arg Asp Leu Arg Ala Cys Pro Cys Met Val Ser Leu Arg Gln Pro Leu
50 55 60

Val Met Leu Cys Ala Val Pro Ala Trp Leu Leu Ala Ser Cys Thr Val
65 70 75 80

His Cys Met Ile Leu His Arg Val Lys His Ala
85 90

<210> 91
<211> 74
<212> PRT
<213> Homo sapien

<400> 91

Met Glu Lys Phe Glu Arg Met Asn Val Lys Ser Phe Phe Phe Phe Phe
1 5 10 15

Phe Glu Thr Gly Ser Leu Ser Val Thr Lys Gln Glu Cys Ser Gly Val
20 25 30

Ile Ile Ala His Cys Ser Leu Asp Leu Pro Gly Ser Ser Asp Pro Pro
35 40 45

Thr Leu Ala Pro Pro Val Ala Gly Thr Thr Gly Val His His His Ser
50 55 60

Trp Leu Ile Ile Ile Leu Phe Leu Tyr Phe
65 70

<210> 92
<211> 92

<212> PRT
<213> Homo sapien

<400> 92

Met Glu His Glu Leu His Pro Thr Ser Gln Ser Cys Gly Ala Arg Ala
1 5 10 15

Thr Ser Ser Ser Val Cys Val Tyr Met Val Glu Leu Ser Leu Cys Asp
20 25 30

Val Ser Leu Ser Arg Ser Pro Cys Phe Gly His Asp Asn Pro Cys Lys
35 40 45

Val Thr Arg Gly Ile Ala Asp Gly Phe Gly Cys Gly Leu Arg Val His
50 55 60

Arg His Val Leu Ala Val Leu Ile Leu Ile Gln Thr Gly Cys Thr Pro
65 70 75 80

Gln Ile Arg Arg Ser Lys Ser Met Ala Ser Val Ala
85 90

<210> 93
<211> 62
<212> PRT
<213> Homo sapien

<400> 93

Met Gly Pro Leu Thr Ala Ala Arg Arg Gly Asp Ser Val Met Asp Gly
1 5 10 15

Trp Cys Asp His Gly Ser Cys Asn Leu Glu Phe Leu Gly Thr Ser Asp
20 25 30

Pro Pro Ala Leu Ala Ser Gln Ser Arg Val Gly Thr Thr Gly Met Arg
35 40 45

Gln His Thr Trp Leu Ile Leu Leu Thr Phe Thr Phe Ser Arg
50 55 60

<210> 94
<211> 148
<212> PRT
<213> Homo sapien

<400> 94

```
Met Leu Gln Lys Gln Asn Thr Arg Ser Gly Gly Gly Glu His Gln Arg
1               5                   10                  15

Glu Gln Pro Met Asp Lys Thr Ala Ser Leu Gly Gly Ser Cys Thr Thr
            20                  25                  30

Pro Arg Ala Pro Pro Thr Phe Thr Val Arg Gly Glu Leu Thr Ala Gln
        35                  40                  45

Lys Val His His Lys Ser Gln Ser Ser Ser His Arg Pro Arg Arg Ala
    50                  55                  60

Ile Pro Gly Gly Gly Thr Lys Arg Lys Lys Arg Asp Ala Gln Ala Ala
65                  70                  75                  80

Asp Ile Ser His Ala Arg Thr Glu His His Gln Asp Thr Arg Gln Asp
                85                  90                  95

Asp Ala Glu Ala Pro His Lys Thr Pro Asn Thr Lys His Pro Arg Thr
            100                 105                 110

Pro Cys Arg His Thr Ala Pro Pro Leu His Pro Pro Glu Gln Met Asn
        115                 120                 125

Arg Gly Gln Ser Asn Thr Arg Arg Asn Glu Asn Asn Leu His Ser Glu
    130                 135                 140

His Asn Ala Ala
145
```

<210> 95
<211> 51
<212> PRT
<213> Homo sapien

<400> 95

```
Met Val Gln Val Leu His Trp Ser Leu Ser Ser Ala Ile Leu Ser Val
1               5                   10                  15

Tyr Val Gln Tyr Leu Pro Gly Asp Pro Ser His Cys Arg Gln Leu Glu
            20                  25                  30

His Ala Ser Met Ile Asn Gln Trp Ala Leu Ile Asn Ser Thr Phe Leu
        35                  40                  45
```

Cys Arg Leu
50

<210> 96
<211> 84
<212> PRT
<213> Homo sapien

<400> 96

Met Arg Gln Ser Ala Thr Leu Arg Ser Ser Asp His Trp Glu Glu Arg
1 5 10 15

Glu Ser Leu Gln Leu Leu Gly Phe Arg Leu Gln Lys Phe Leu Ala Ala
20 25 30

Phe Ala His Trp Arg Gly Gly Glu Asp Lys Ser Ile Arg Asn Pro Met
35 40 45

Phe Pro Ser Ser Pro Thr Glu Arg Thr Lys Glu Val Phe Thr Arg Cys
50 55 60

Gly Thr Phe Leu Gln Leu Leu Asp Ala Asp Lys Pro Gln Ser Arg Leu
65 70 75 80

Phe Trp Leu Gln

<210> 97
<211> 72
<212> PRT
<213> Homo sapien

<400> 97

Met Lys Gln Trp Lys Ile Ser Ile Ala Gln Leu Asp Asp Leu Thr Lys
1 5 10 15

Glu Ile Ser Arg Gln Cys Gln Arg Cys Tyr Leu Asp Ser Ser Ser Pro
20 25 30

Tyr Ser Lys Arg Gln Lys Glu Lys Gly Lys Gln Asp Lys Lys Leu Phe
35 40 45

Asp Ile Lys Glu Pro Gln Leu Phe Gly Phe Glu Lys Tyr Phe Phe Ser
50 55 60

Phe Leu Thr Ser Pro Asp Ser Glu

65 70

<210> 98
<211> 40
<212> PRT
<213> Homo sapien

<400> 98

Met Gly Thr Arg Tyr Tyr Ile Leu Glu Phe Val Leu Arg Arg His Lys
1 5 10 15

Leu Asn Ser Arg Ser Leu Cys Pro Lys Phe His Arg Leu Lys Lys Arg
20 25 30

Ser Ser Asn Tyr Arg Ser Gly Tyr
35 40

<210> 99
<211> 87
<212> PRT
<213> Homo sapien

<400> 99

Met Phe Ser Thr Ser Ser Gln Val Cys Ala Leu Cys Pro Phe Ser Gly
1 5 10 15

Ser Leu Glu Leu Pro Pro Ser Leu His Pro Asp Ser Phe Ala Ile Met
20 25 30

Cys Leu Ile Ser Cys Glu Phe Thr Gly Glu Ala Ile Ser Gln Ile Asn
35 40 45

Gly Cys Lys Cys Ser Lys Lys Lys Lys Thr Lys Lys Lys Ala Gly Gly
50 55 60

Asn Arg Gly Gln Ser Leu Ser Pro Gly Gly His Cys Phe Pro Pro Gln
65 70 75 80

Phe Asn Pro His Lys Pro Pro
85

<210> 100
<211> 31
<212> PRT
<213> Homo sapien

<400> 100

Met Ser Asn Ser His Thr Glu Gln Ala Thr Phe Leu Ser Lys Val Cys
1 5 10 15

Gly Ala Gly Arg Ala Val Gly Ala Leu Asn Ala Gly Leu Asn Arg
20 25 30

<210> 101
<211> 69
<212> PRT
<213> Homo sapien

<400> 101

Met Leu Arg Asn Cys Gly Gly Ile Gly Ala Gly Asn Lys Phe Pro Pro
1 5 10 15

Gly Thr Ala Leu Ala Pro Asp Thr Pro Ser Leu Phe Phe Phe Phe Phe
20 25 30

Phe Phe Leu Glu Thr Met Thr Thr Ala Ala Ala Ile Leu Leu Pro Ile
35 40 45

Ser His Glu Pro Arg Leu Pro Tyr Thr Met Thr Phe His Pro His Asn
50 55 60

Arg Leu Thr Gln Pro
65

<210> 102
<211> 91
<212> PRT
<213> Homo sapien

<400> 102

Met Phe Cys Val Phe Leu Lys Ser Glu Cys Val Phe Tyr His Cys Ser
1 5 10 15

Val Asn Ala Asn Trp Val Lys Phe Val Asp Ser Gln Ile Tyr Ile Leu
20 25 30

Thr His Leu Phe Val Pro Phe Phe Leu Ser Val Ile Glu Gln Glu Val
35 40 45

Leu Lys Ser Pro Ile Thr Ser Ile Ser Leu Thr Leu Pro Phe Phe Ser
50 55 60

Leu Trp Ile Leu Asn Phe Ser Ile Tyr Phe Val Tyr Phe Glu Gly His
65 70 75 80

Ile His Leu Leu Ser Ser Cys Ile Leu Met Asn
85 90

<210> 103
<211> 38
<212> PRT
<213> Homo sapien

<400> 103

Gln Pro Gly Gln His Gly Glu Thr Pro Ser Pro Pro Lys Asp Ala Lys
1 5 10 15

Thr Ser Gln Ala Trp Arg Arg Ala Pro Ala Val Pro Gly Thr Arg Gln
20 25 30

Ala Glu Ala Gly Glu Ser
35

<210> 104
<211> 107
<212> PRT
<213> Homo sapien

<400> 104

Met Asn Tyr Ser Leu Thr Ser Arg Thr Val Glu Asp Arg Gly Gln Lys
1 5 10 15

Gln Ala Ser Lys Arg Ser Gln Tyr Gly Gly Val His Ala Trp His Thr
20 25 30

Trp Leu Ser Glu Ser Asp Val Cys Leu Cys Val Cys Asp Glu Asp Ser
35 40 45

Ser Glu Trp Asn Gly Gln Arg Val Thr Gly Lys Phe Cys Arg Glu Glu
50 55 60

Asn Glu Arg Leu Leu Ile Leu Lys Gln Ser Phe Ala Leu Leu Trp Ser
65 70 75 80

Tyr Thr Thr Val Asn Leu Pro Ile Leu Ser Ser Gln Ile Pro Thr Arg
85 90 95

Lys Pro Val Ile Asn Leu Trp Ile Asn Phe His

100 105

<210> 105
<211> 822
<212> PRT
<213> Homo sapien

<400> 105

Met Asn Thr Ala Asp Gln Ala Arg Val Gly Pro Ala Asp Asp Gly Pro
1 5 10 15

Ala Pro Ser Gly Glu Glu Glu Gly Glu Gly Gly Gly Glu Ala Gly Gly
20 25 30

Lys Glu Pro Ala Ala Asp Ala Ala Pro Gly Pro Ser Ala Ala Phe Arg
35 40 45

Leu Met Val Thr Arg Arg Glu Pro Ala Val Lys Leu Gln Tyr Ala Val
50 55 60

Ser Gly Leu Glu Pro Leu Ala Trp Ser Glu Asp His Arg Val Ser Val
65 70 75 80

Ser Thr Ala Arg Ser Ile Ala Val Leu Glu Leu Ile Cys Asp Val His
85 90 95

Asn Pro Gly Gln Asp Leu Val Ile His Arg Thr Ser Val Pro Ala Pro
100 105 110

Leu Asn Ser Cys Leu Leu Lys Val Gly Ser Lys Thr Glu Val Ala Glu
115 120 125

Cys Lys Glu Lys Phe Ala Ala Ser Lys Asp Pro Thr Val Ser Gln Thr
130 135 140

Phe Met Leu Asp Arg Val Phe Asn Pro Glu Gly Lys Ala Leu Pro Pro
145 150 155 160

Met Arg Gly Phe Lys Tyr Thr Ser Trp Ser Pro Met Gly Cys Asp Ala
165 170 175

Asn Gly Arg Cys Leu Leu Ala Ala Leu Thr Met Asp Asn Arg Leu Thr
180 185 190

Ile Gln Ala Asn Leu Asn Arg Leu Gln Trp Val Gln Leu Val Asp Leu

```
        195                 200                 205

Thr Glu Ile Tyr Gly Glu Arg Leu Tyr Glu Thr Ser Tyr Arg Leu Ser
    210                 215                 220

Lys Asn Glu Ala Pro Glu Gly Asn Leu Gly Asp Phe Ala Glu Phe Gln
225                 230                 235                 240

Arg Arg His Ser Met Gln Thr Pro Val Arg Met Glu Trp Ser Gly Ile
                245                 250                 255

Cys Thr Thr Gln Gln Val Lys His Asn Asn Glu Cys Arg Asp Val Gly
            260                 265                 270

Ser Val Leu Leu Ala Val Leu Phe Glu Asn Gly Asn Ile Ala Val Trp
        275                 280                 285

Gln Phe Gln Leu Pro Phe Val Gly Lys Glu Ser Ile Ser Ser Cys Asn
    290                 295                 300

Thr Ile Glu Ser Gly Ile Thr Ser Pro Ser Val Leu Phe Trp Trp Glu
305                 310                 315                 320

Tyr Glu His Asn Asn Arg Lys Met Ser Gly Leu Ile Val Gly Ser Ala
                325                 330                 335

Phe Gly Pro Ile Lys Ile Leu Pro Val Asn Leu Lys Ala Val Lys Gly
            340                 345                 350

Tyr Phe Thr Leu Arg Gln Pro Val Ile Leu Trp Lys Glu Met Asp Gln
        355                 360                 365

Leu Pro Val His Ser Ile Lys Cys Val Pro Leu Tyr His Pro Tyr Gln
    370                 375                 380

Lys Cys Ser Cys Ser Leu Val Val Ala Ala Arg Gly Ser Tyr Val Phe
385                 390                 395                 400

Trp Cys Leu Leu Leu Ile Ser Lys Ala Gly Leu Asn Val His Asn Ser
                405                 410                 415

His Val Thr Gly Leu His Ser Leu Pro Ile Val Ser Met Thr Ala Asp
            420                 425                 430
```

Lys Gln Asn Gly Thr Val Tyr Thr Cys Ser Ser Asp Gly Lys Val Arg
435 440 445

Gln Leu Ile Pro Ile Phe Thr Asp Val Ala Leu Lys Phe Glu His Gln
450 455 460

Leu Ile Lys Leu Ser Asp Val Phe Gly Ser Val Arg Thr His Gly Ile
465 470 475 480

Ala Val Ser Pro Cys Gly Ala Tyr Leu Ala Ile Ile Thr Thr Glu Gly
485 490 495

Met Ile Asn Gly Leu His Pro Val Asn Lys Asn Tyr Gln Val Gln Phe
500 505 510

Val Thr Leu Lys Thr Phe Glu Glu Ala Ala Ala Gln Leu Leu Glu Ser
515 520 525

Ser Val Gln Asn Leu Phe Lys Gln Val Asp Leu Ile Asp Leu Val Arg
530 535 540

Trp Lys Ile Leu Lys Asp Lys His Ile Pro Gln Phe Leu Gln Glu Ala
545 550 555 560

Leu Glu Lys Lys Ile Glu Ser Ser Gly Val Thr Tyr Phe Trp Arg Phe
565 570 575

Lys Leu Phe Leu Leu Arg Ile Leu Tyr Gln Ser Met Gln Lys Thr Pro
580 585 590

Ser Glu Ala Leu Trp Lys Pro Thr His Glu Asp Ser Lys Ile Leu Leu
595 600 605

Val Asp Ser Pro Gly Met Gly Asn Ala Asp Asp Glu Gln Gln Glu Glu
610 615 620

Gly Thr Ser Ser Lys Gln Val Val Lys Gln Gly Leu Gln Glu Arg Ser
625 630 635 640

Lys Glu Gly Asp Val Glu Glu Pro Thr Asp Asp Ser Leu Pro Thr Thr
645 650 655

Gly Asp Ala Gly Gly Arg Glu Pro Met Glu Glu Lys Leu Leu Glu Ile
660 665 670

Gln Gly Lys Ile Glu Ala Val Glu Met His Leu Thr Arg Glu His Met
675 680 685

Lys Arg Val Leu Gly Glu Val Tyr Leu His Thr Trp Ile Thr Glu Asn
690 695 700

Thr Ser Ile Pro Thr Arg Gly Leu Cys Asn Phe Leu Met Ser Asp Glu
705 710 715 720

Glu Tyr Asp Asp Arg Thr Ala Arg Val Leu Ile Gly His Ile Ser Lys
725 730 735

Lys Met Asn Lys Gln Thr Phe Pro Glu His Cys Ser Leu Cys Lys Glu
740 745 750

Ile Leu Pro Phe Thr Asp Arg Lys Gln Ala Val Cys Ser Asn Gly His
755 760 765

Ile Trp Leu Arg Cys Phe Leu Thr Tyr Gln Ser Cys Gln Ser Leu Ile
770 775 780

Tyr Arg Arg Cys Leu Leu His Asp Ser Ile Ala Arg His Pro Ala Pro
785 790 795 800

Glu Asp Pro Asp Trp Ile Lys Arg Leu Leu Gln Ser Pro Cys Pro Phe
805 810 815

Cys Asp Ser Pro Val Phe
820

<210> 106
<211> 52
<212> PRT
<213> Homo sapien

<400> 106

Met Asn Tyr Val Leu Asn Glu Trp Leu Ser Leu Pro Cys Lys Pro His
1 5 10 15

Ala Thr Gly Ser Leu Phe Arg Trp Leu Thr Thr Ala Pro Gln Ala Cys
20 25 30

Trp Lys Asp Arg Ser Pro Lys Pro Ser Leu Leu Ser Thr Gln Ala Trp
35 40 45

Val Ser Trp Ser
50

<210> 107
<211> 82
<212> PRT
<213> Homo sapien

<400> 107

Met Leu Asn Thr Cys Arg Val Ile Leu Val Val Phe Ser Gln Pro Phe
1 5 10 15

Ile Lys Phe Leu Val Thr Ser Val Met Met Thr Phe His Thr Pro Ile
20 25 30

Thr Ser Lys Ala Phe Leu His Leu Ala Asp Pro Ser Tyr Gly Pro Ala
35 40 45

Val Ser His Ala Val Thr Thr Ser Gly Thr Asp Leu Thr Ala Leu Arg
50 55 60

Ala Ser Ser Ser Leu Ala Gly Arg Thr Ser Ala Ala Ser Ser Ile Thr
65 70 75 80

Lys Gly

<210> 108
<211> 63
<212> PRT
<213> Homo sapien

<400> 108

Met Arg Val Ser Gly Thr Cys Trp Asp Lys Cys Glu Ala Ser Val Trp
1 5 10 15

Ala Val Arg Tyr Gly Glu Cys Leu Ser Leu Arg Ser Lys Glu Leu Trp
20 25 30

Ala Gly Pro Trp Arg Trp Arg Arg Val Pro Val Val Ser Ala Lys Ser
35 40 45

Gly Gly Arg Lys Trp Glu Asp His Leu Ser Pro Gly Ile Arg Gly
50 55 60

<210> 109
<211> 51
<212> PRT
<213> Homo sapien

<400> 109

Val Cys Gly Gly Ser Arg Gln Arg Gln Gly Leu Ala Pro Leu Ser Arg
1 5 10 15

Leu Glu Cys Phe Gly Val Met Thr Ala His Val Asn Leu Glu Phe Leu
20 25 30

Gly Ser Gly Asp Pro Pro Thr Ser Ala Ser Ala Leu Ala Glu Thr Thr
35 40 45

Gly Thr Arg
50

<210> 110
<211> 141
<212> PRT
<213> Homo sapien

<400> 110

Met Ile Leu Leu Ser Arg His Asn Ser Gln Gly Asn Thr Thr Thr His
1 5 10 15

His Asn Lys Asn Thr Lys Thr Arg Gly Gly Asp Thr Pro Gly Thr Thr
20 25 30

Gly Trp Ile Pro Gly Arg Arg Thr Arg Ser Pro Arg Arg Gln Asn Phe
35 40 45

Pro Thr Lys Thr Ile Gly Asp Lys Thr Ala Lys Glu Ala Arg Glu Thr
50 55 60

Arg Gly Asn Lys Arg Lys Lys Asp Thr Glu Arg Arg Lys Gly Ala Arg
65 70 75 80

Ser Thr Arg Thr Arg Asp Glu Glu Gly Gly Gly Arg Glu Glu Glu Arg
85 90 95

Gly Arg Gly Gly Arg Glu Arg Arg Gln Glu Gly Glu Arg Gly Ile Glu
100 105 110

Thr Gly Gly Glu Glu Glu Arg Lys Arg Gly Gly Arg Gly Arg Gly Gly

115 120 125

Glu Arg Arg Gly Gly Lys Lys Glu Asp Gly Gly Pro Glu
130 135 140

<210> 111
<211> 99
<212> PRT
<213> Homo sapien

<400> 111

Met Gly Arg Trp Glu Glu Ser Gln Ser Thr Gly Gln Gly Glu Asp Ser
1 5 10 15

Gly Ser His Gly Val Ser Pro Thr Ala Ser Ala Pro Leu Cys Cys Trp
20 25 30

Arg Gly Pro Glu Pro His Tyr Ser Leu Tyr Glu Asp Gln Ser Val Phe
35 40 45

Gly Arg Trp Arg Leu Ala His Gly Arg Thr Pro Ser Gly Gly Gly Ser
50 55 60

Ser Val Asn Pro Arg Asn Phe Lys Glu Pro His Ser Val Ser Leu Met
65 70 75 80

Thr Ser His Leu Gln Ile Arg Lys Leu Trp Ile Pro Arg Gly Ser Phe
85 90 95

Gly Ser Ile

<210> 112
<211> 105
<212> PRT
<213> Homo sapien

<400> 112

Gly Ala Gly Gly Tyr Ala Asp Asn Asp Ile Gly Ala Val Ser Thr Thr
1 5 10 15

Gly His Gly Glu Ser Ile Leu Lys Val Asn Leu Ala Arg Leu Thr Leu
20 25 30

Phe His Ile Glu Gln Gly Lys Thr Val Glu Glu Ala Ala Asp Leu Ser
35 40 45

Leu Gly Tyr Met Lys Ser Arg Val Lys Gly Leu Gly Gly Leu Ile Val
50 55 60

Val Ser Lys Thr Gly Asp Trp Val Ala Lys Trp Thr Ser Thr Ser Met
65 70 75 80

Pro Trp Ala Ala Ala Lys Asp Gly Lys Leu His Phe Gly Ile Asp Pro
85 90 95

Asp Asp Thr Thr Ile Thr Asp Leu Pro
100 105

<210> 113
<211> 42
<212> PRT
<213> Homo sapien

<400> 113

Met Ala Thr Pro Pro Ala Lys Cys Leu Ser Gln Asp Leu Asp Ser Ser
1 5 10 15

Pro Trp Asp Pro His Ala Arg Glu Ala Asp Cys Ser Ala Pro Thr Gly
20 25 30

Ser Leu His Glu Val Val Pro Gln His Cys
35 40

<210> 114
<211> 51
<212> PRT
<213> Homo sapien

<400> 114

Met Leu Leu Ser Tyr Ile Ser Gly Arg Phe Leu Ser Thr Arg Lys Glu
1 5 10 15

Asn Thr Gly Leu Ala Lys Gln Gly Pro Leu Phe Gly Ile Ile Phe Val
20 25 30

Pro Asn Lys Gln Ser Arg Gly Trp Val Cys Trp Leu Val Lys Glu Leu
35 40 45

Leu Arg Phe
50

<210> 115
<211> 118
<212> PRT
<213> Homo sapien

<400> 115

Met Asp Glu Arg Arg Pro Gly Arg Tyr Leu Gly Leu Pro Glu Tyr Thr
1 5 10 15

Lys Phe Arg Glu Pro Thr Phe Thr Pro Asp Cys Ala Trp Ser Lys Pro
20 25 30

Glu Ser Ser Leu Pro Arg Gly Leu Phe Gln Pro Ile Pro Leu Phe Trp
35 40 45

Lys Val Ile Leu Gly Ile Glu Thr Glu Asn Trp Asp Lys Gly Ser Leu
50 55 60

Arg Lys Thr Lys Thr Asn Asn Glu Thr Gly Asp Met Leu Phe Ser Leu
65 70 75 80

Asn Pro Ser Gln Ile Cys Cys Leu Ala Leu Thr His Val Glu Ile Cys
85 90 95

Lys Leu Cys Gln Asp Phe Pro Val His Gly Gly Glu Ser His Val Gly
100 105 110

Lys Lys Lys Phe Thr Val
115

<210> 116
<211> 87
<212> PRT
<213> Homo sapien

<400> 116

Met Leu Glu Arg Arg Ser Val Met Asp Trp Ser Arg Arg Gly Leu Trp
1 5 10 15

Glu Pro Gly Leu Gln Cys Gly Leu Pro Arg Pro Pro Gly Pro Ser Ala
20 25 30

Ser Ser Leu Arg Gln Pro Ser Gln Gly Trp Pro Ala Arg Thr Asp Val
35 40 45

```
Thr Met Pro Arg Ala Pro Ala Pro His Thr Ala Glu Leu Met Met Val
    50                  55                  60

Met Gly Gly Ser Gly Ala Gly Ala Gly Glu Gln Asp Glu Gln Glu Cys
65                  70                  75                  80

Asn Asn Gln Asp Asp Pro Glu
                85
```

<210> 117
<211> 72
<212> PRT
<213> Homo sapien

<400> 117

```
Met His Val Pro Thr Glu Arg Glu Tyr Ala Cys Val Cys Thr Thr Asn
1               5                   10                  15

Thr Ser Cys Cys Ala Gly Ala Gly Ser Ser Gly Asn Ala Arg Gly Glu
            20                  25                  30

His Ala Leu Leu Val Ile His Ile His Ser Tyr Ala Val His Thr Gln
        35                  40                  45

His Pro Pro Arg Ala Cys Leu Pro Asn Arg Trp Leu Asn Phe Leu Leu
    50                  55                  60

Ser Tyr Arg Arg Pro Asp Pro Thr
65                  70
```

<210> 118
<211> 48
<212> PRT
<213> Homo sapien

<400> 118

```
Met Asn Pro Arg Ile Asn Thr Leu Asp Val Leu Leu Leu Cys His Val
1               5                   10                  15

Asn Arg Gly Leu Arg Ala Val Phe His Leu Val Pro Phe Ser Glu Asp
            20                  25                  30

Gln Ile Pro Arg Leu Gln Ser Met Gln Gly Leu His Arg Trp Leu Leu
        35                  40                  45
```

<210> 119

<211> 19
<212> PRT
<213> Homo sapien

<400> 119

Met Thr Trp Thr Asn Arg Lys Tyr Ser Phe Asn Leu Phe Leu Leu Leu
1 5 10 15

Phe Asn Leu

<210> 120
<211> 60
<212> PRT
<213> Homo sapien

<400> 120

Met Thr Phe Gly Val Pro Asn Ser Val Ser Thr Leu Thr Ser Lys Lys
1 5 10 15

Lys Lys Arg Lys Lys Lys Lys Gly Arg Gly Val Pro Trp Gly Asn Ser
20 25 30

Cys Pro Gly Gly Gly Ile Val Phe Pro Val Pro Ile Pro Pro Ile Phe
35 40 45

His Asn Asn Gly Glu Pro Gly Gln Lys Arg Lys Thr
50 55 60

<210> 121
<211> 147
<212> PRT
<213> Homo sapien

<400> 121

Met Leu Leu Glu Arg Arg His Cys Asp Gly Cys Val Val Ala Pro Arg
1 5 10 15

Leu Cys Val Lys Arg Glu Ala Glu Gly Asp Val Ser Pro Asp Ile Ser
20 25 30

Lys Val Trp Val Gly Pro Leu Val Pro Glu Ile Leu Leu Gly Gly Met
35 40 45

Gly Pro Ala Leu Ser Gly Thr Lys Ile Arg Ala Arg Lys Arg Cys Pro
50 55 60

Ser Pro Ile Leu Ser Ile Leu Phe Met Ala Glu Lys Ile Ser Ala Gly
65 70 75 80

Cys Gln His Val Pro Met Pro Val Glu Asp Met Pro Thr Ser Pro Leu
85 90 95

Pro Arg Glu Gln Asp Leu Gly Leu Gly Gln Val Glu Lys Ile Pro Asp
100 105 110

Phe Phe Ser Thr Val Phe Val Leu Met Val Tyr Phe Tyr Trp Leu Leu
115 120 125

Tyr Cys Leu Gly Gln Val Val Val Ala Phe Leu Ile Tyr Trp Gly Thr
130 135 140

Phe Leu Ile
145

<210> 122
<211> 121
<212> PRT
<213> Homo sapien

<400> 122

Met Val Arg Ile Leu Ala Asn Gly Glu Ile Val Gln Asp Asp Asp Pro
1 5 10 15

Arg Val Arg Thr Thr Thr Gln Pro Pro Arg Gly Ser Ile Pro Arg Gln
20 25 30

Ser Phe Phe Asn Arg Gly His Gly Ala Pro Pro Gly Gly Pro Gly Pro
35 40 45

Arg Gln Gln Gln Ala Gly Ala Arg Leu Gly Ala Ala Gln Ser Pro Phe
50 55 60

Asn Asp Leu Asn Arg Gln Leu Val Asn Met Gly Phe Pro Gln Trp His
65 70 75 80

Leu Gly Asn His Ala Val Glu Pro Val Thr Ser Ile Leu Leu Leu Phe
85 90 95

Leu Leu Met Met Leu Gly Val Arg Gly Leu Leu Leu Val Gly Leu Val
100 105 110

Tyr Leu Val Ser His Leu Ser Gln Arg
115 120

<210> 123
<211> 129
<212> PRT
<213> Homo sapien

<400> 123

Met Glu Ala Arg Arg His Ala Leu Gly Gly Ser Val Leu Trp Gln Ser
1 5 10 15

Gln Val Leu Phe Asn Phe Val Gln Arg Lys Gly Glu Pro Gly Phe Gly
20 25 30

Ile Ser Val Val Arg Glu Arg Arg Val His Ser Asn His Gly Cys Pro
35 40 45

Val Leu Ile Gln Ala Gly Ile Trp Ser Met Met Ser Val Gly Arg Ala
50 55 60

Arg Arg Ala Arg Arg Thr Ala Ala Ser Tyr Pro Gly Pro Val Arg Ala
65 70 75 80

Tyr Leu His His Ala Arg Gly Gly Gln Glu Pro Pro Pro Ala Val Pro
85 90 95

Ala Arg Ala Gly Ser Ile Thr Leu Ser Pro Leu Glu Met Ile Arg Gly
100 105 110

Pro Ser Pro Tyr Glu Ser Ile Ser Tyr Leu Ser Arg Gly Val Phe Leu
115 120 125

Leu

<210> 124
<211> 74
<212> PRT
<213> Homo sapien

<400> 124

Met Lys Ile Tyr Leu Ser Ser Leu Ile Leu Gln Val Thr Ile Ile Leu
1 5 10 15

Asn Pro Ile Lys Ser Trp Ala Val Ala Arg Phe Phe Phe Phe Phe Arg
20 25 30

Gly Gly Pro Lys Glu Ala Ser Gln Gly Arg Leu Pro Gly Leu Cys Pro
35 40 45

Pro Pro Leu Ala Phe Ala Leu Cys Ser Gln Cys Ser Ser Ser Lys Arg
50 55 60

Ala Ser Leu Ser Pro Gln Pro Pro Pro Cys
65 70

<210> 125
<211> 94
<212> PRT
<213> Homo sapien

<400> 125

Met His Ser Gly Trp Glu Trp Trp Leu Met Pro Val Ile Pro Ala Val
1 5 10 15

Cys Gly Gly Pro Gln Val Asp Arg Leu Phe Asp Ala Gln Ala Val Arg
20 25 30

Asp Gln Pro Gly Val Thr Met Gly Gly Thr Pro Asn Leu Tyr Gln Lys
35 40 45

Lys Lys Lys Asn Thr Lys Val Val Trp Val Arg Gly Arg Met Pro Val
50 55 60

Val Pro Lys Phe Pro Ala Thr Leu Leu Gly Arg Leu Arg Gln Lys Gly
65 70 75 80

Ser Pro Glu Pro Arg Glu Gly Pro Arg Leu Ala Val Ser Pro
85 90

<210> 126
<211> 114
<212> PRT
<213> Homo sapien

<400> 126

Met Val Ser Leu Trp Val Glu Asp Thr Phe Leu Ser Pro Gly Phe Gly
1 5 10 15

Phe Ala His Val Ala Cys Ser Gly Leu Gly Met Lys Gln Lys Arg Lys

20 25 30

Ala Ala Ser Ser Glu Pro Thr Ser Glu Val Ala Leu Gly Gly Ser Ala
35 40 45

Gly Pro Val Arg Ser His Leu His Pro Glu Gly Leu Leu Trp Cys Ser
50 55 60

Arg Cys Phe Phe Ser Leu Arg Pro Lys Gly Thr Glu Pro Pro Gly Arg
65 70 75 80

Ser Ala Gly Leu Gln Gly Ala Thr Glu Arg Ser Gly Trp Thr Ser Val
85 90 95

Gln Ala Gln Ala Gln Ala Cys Glu Asn Leu Val Pro Ala Ala Val Ala
100 105 110

Asp Gly

<210> 127
<211> 27
<212> PRT
<213> Homo sapien

<400> 127

Met Asn Ser Phe Tyr Cys Lys Gln Thr Ser Lys Leu Ile Ser Pro Pro
1 5 10 15

Thr Phe Phe Arg Lys Lys Lys Lys Ser Ala Gly
20 25

<210> 128
<211> 59
<212> PRT
<213> Homo sapien

<400> 128

Met Tyr Ser Tyr Asn Gly Ile Leu Phe Asp Asn Lys Asn Lys Trp Ser
1 5 10 15

Ala Ser Thr Cys Tyr Asn Lys Lys Lys Lys Lys Lys Lys Thr Leu Gly
20 25 30

Leu Ser His Gly Ser Tyr Leu Phe Pro Cys Phe Asp Ile Phe Phe Pro
35 40 45

Leu Pro Ile Ser Thr Gln Ile Leu Thr Gln Ile
50 55

<210> 129
<211> 110
<212> PRT
<213> Homo sapien

<400> 129

Met Lys Pro Arg Thr Leu Gly Pro Ser Leu Lys Ile Pro Ala Pro Gly
1 5 10 15

Cys Gly Lys Leu His Ala Pro Ser Phe Ser Ser Thr Leu Met Leu Pro
20 25 30

Gly Val Cys Ser Tyr Arg Thr Pro Thr Pro Ala Thr Leu Gln Glu Asp
35 40 45

Gly Lys Pro Gln Thr Pro Leu His Ser Lys Glu Ser His Gln Ala Thr
50 55 60

Arg Gly Ile Gln Leu Ala Pro Ser Leu His Met Val Gly Gly Asp Gln
65 70 75 80

Arg His Gly Thr Asp Ala Gly Cys Ala Leu Trp Pro Pro Asn Leu Ile
85 90 95

Leu Val Thr Ser Pro Phe Ala Thr Met Arg Ala Gln Glu Met
100 105 110

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 November 2002 (21.11.2002)

PCT

(10) International Publication Number
WO 02/092785 A3

(51) International Patent Classification[7]: C12N 15/12, 15/11, 15/00, C12Q 1/68, C12P 21/02, A61K 45/00, 48/00

(21) International Application Number: PCT/US02/22271

(22) International Filing Date: 13 February 2002 (13.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/268,290	13 February 2001 (13.02.2001)	US
60/268,834	15 February 2001 (15.02.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/268,290 (CIP)
Filed on	13 February 2001 (13.02.2001)
US	60/268,834 (CIP)
Filed on	15 February 2001 (15.02.2001)

(71) Applicant *(for all designated States except US)*: **DIADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SALCEDA, Susana** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **MACINA, Roberto, A.** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). **HU, Ping** [US/US]; 108 Palmer Street, San Ramon, CA 94583 (US). **RECIPON, Herve** [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). **KARRA, Kalpana** [IN/US]; 6511 Rainbow Drive, San Jose, CA 95129 (US). **CAFFERKEY, Robert** [IE/US]; 849 West Orange Avenue, Apartment 1030, South San Francisco, CA 94081 (US). **SUN, Yongming** [US/US]; Apartment 260, 869 S. Winchester Boulevard, San Jose, CA 95128 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrel P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
11 December 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/092785 A3

(54) Title: COMPOSITION AND METHODS RELATING TO OVARIAN SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of the compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovary tissue, identifying ovary tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/22271

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 15/12, 15/11, 15/00; C12Q 1/68; C12P 21/02; A61K 45/00, 48/00
US CL : 536/23.5, 23.1; 435/6, 320.1, 325, 69.1; 514/44

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.5, 23.1; 435/6, 320.1, 325, 69.1; 514/44

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Compugen, SEQ ID NO: 77, nucleic acid and amino acid databases

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CRABB, J.W. et al, Cloning of the cDNAs encoding the cellular retinaldehyde-binding protein from bovine and human retina and comparison of the protein structures, Journal of Biological Chemistry, December 1988, Vol. 263, No. 35, pages 18688-18692, see especially Figure 1, positions 221-250 (complement) and alignment attached to reference.	1, 2, 4, 5, 7, and 8

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
17 July 2003 (17.07.2003)

Date of mailing of the international search report
04 AUG 2003

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer
James Martinell

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/22271

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-9, 14/1, 15/1, and 17 as they pertain to SEQ ID NO: 77

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/22271

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-9, 14/1, 15/1, and 17, drawn to nucleic acids, nucleic acid molecular hybridization assays, vectors, host cells, methods for producing polypeptides, kits, and polynucleotide vaccines.

Group II, claim(s) 10, 11, 15/11, and 17, drawn to polypeptides, kits, and polypeptide vaccines.

Group III, claim(s) 12, 13, 14/11, and 16, drawn to antibodies, antibody assays, and methods of treatment using antibodies.

The inventions listed as Groups I-III do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

The nucleic acids, vectors, host cells, kits, and vaccines of Group I are materially different from the polypeptides, kits, and vaccines of Group II and the antibodies of Group III. The methods of Group I may be practiced without the polypeptides, kits, and vaccines of Group II and without the antibodies of Group III. The methods of Groups I and III may be practiced independently of one another. The polypeptides, kits, and vaccines of Group II are materially different from the \abs of Group III and are not needed to practice the methods of Group III.

Each of the Groups mentions or requires the use of a large number of separate and unrelated nucleic acids and/or polypeptides. No matter which additional Group applicant elects, applicant is further required to select for search on SEQ ID NO within the Group for search. Any additional SEQ ID NO to be searched requires on additional search fee per SEQ ID NO. In the absence of payment of additional search fee(s) the first mentioned SEQ ID NO in Group I will be searched. Should applicant pay fee(s) for additional Groups to be searched, the first mentioned SEQ ID NO within the selected Group will be searched unless applicant directs otherwise.